

# AJNR

## AMERICAN JOURNAL OF NEURORADIOLOGY

Official Journal ASNR • ASFNR • ASHNR • ASPNR • ASSR

JULY 2025 | VOLUME 46 | NUMBER 7 | WWW.AJNR.ORG

THE JOURNAL OF DIAGNOSTIC AND INTERVENTIONAL NEURORADIOLOGY

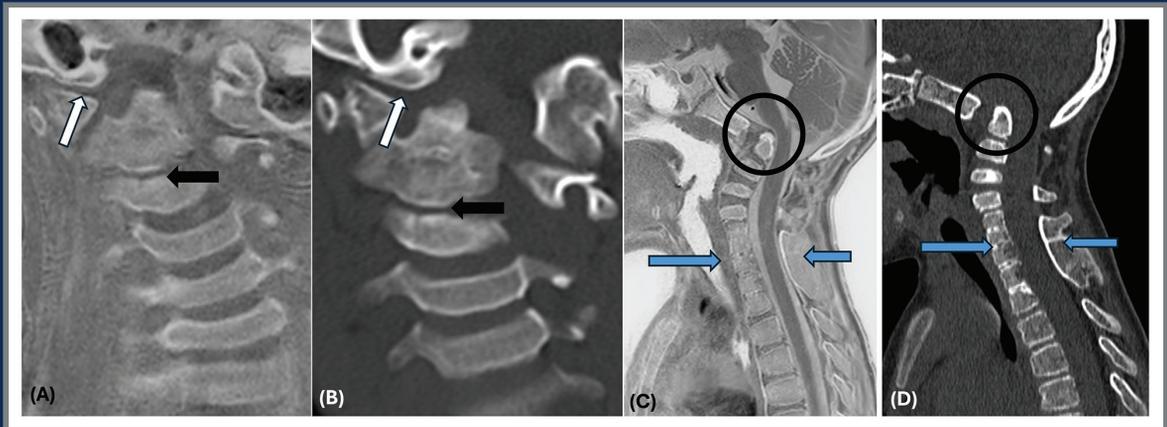
CPT codes for MRI safety

CBF changes: biomarker of antiamyloid therapy response

Peritumoral halo around vestibular schwannoma post SRS

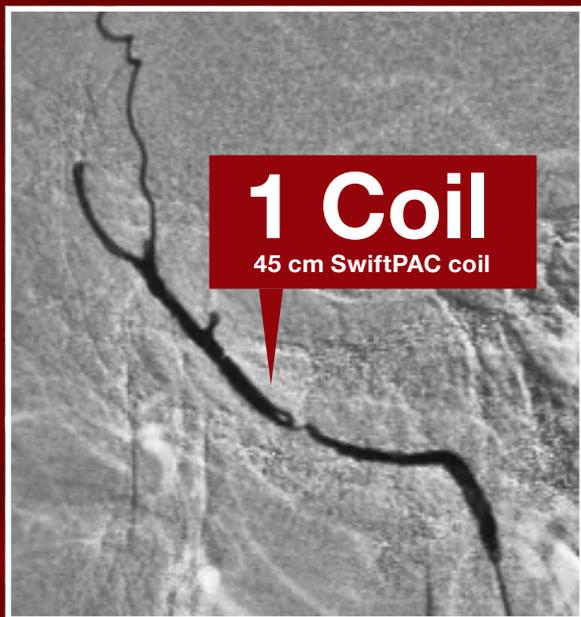
ADC for grading pediatric brain tumors

Image-guided biopsy impacts management of spondylodiscitis



# Do More, With Less

Designed as  
Liquid Metal



Dr. Angelos Konstas  
Huntington Memorial Hospital, CA

**swiftPAC**<sup>™</sup>  
Packing Coil

Optimized for Intracranial  
Flow Disruption<sup>a</sup>



Dr. Alan North  
St. Dominic Jackson Memorial Hospital, MS

**PC400**<sup>™</sup>  
Second Generation



**Learn More Here**

For the complete Penumbra IFU Summary Statements and more, please scan QR code or visit: [peninc.info/neuro-emb](https://peninc.info/neuro-emb)

a. Berge J, Gariel F, Marnat G, Dousset V. PC400 volumetric coils minimize radiation, reduce procedure time and optimize packing density during endovascular treatment in medium sized cerebral aneurysms. *J Neuroradiol*. 2016 Feb;43(1):37-42. doi:10.1016/j.neurad.2015.10.002.

Images used with permission. Consents on file at Penumbra, Inc. Photographs taken by and on file at Penumbra, Inc. Individual results may vary depending on patient-specific attributes and other factors.

Caution: Federal (USA) law restricts these devices to sale by or on the order of a physician. Prior to use, please refer to the Instructions for Use (IFU) for complete product indications, contraindications, warnings, precautions, potential adverse events, and detailed instructions for use. Please contact your local Penumbra representative for more information.

Copyright ©2024 Penumbra, Inc. All rights reserved. The Penumbra P logo, Swift, SwiftPAC, Penumbra Coil 400, and PC400 are registered trademarks or trademarks of Penumbra, Inc. in the USA and other countries. 31204, Rev. A 11/24 USA

## Course Speakers:

### Center for Advance Practice

Rachel Turner, MSN, CRNA, NE-BC

### MRI Supervisor

Annie Pindell, MHA, R.T.(R) (MR) (ARRT)

### Neurology

Raja B. Khan, MD

### Neuroradiology

Asim K. Bag, MBBS, MD, EDiNR

Thierry A. G. Huisman, MD, PD, FICIS, FACR

Soniya N. Pinto, MBBS, MS

Tina Young Poussaint, MD, FACR

Sumit Pruthi, MBBS, MD

Noah D. Sabin, MD, JD

Kathleen M. Schmainda, PhD

Benita Tamrazi, MD

Arastoo Vossough, MD, PhD

### Neuro Oncology

Emily Hanzlink, MD

Giles W. Robinson, MD

### Neuro-Surgery

Paul D. Klimo Jr., MD, MPH

Nir Shimony, MD

### Nuclear Medicine

Barry L. Shulkin, MD, MBA

### Oncology

Aditi Bagchi, MD, PhD

Kelsey Bertrand, MD

Patrick K. Campbell, MD, PhD

Melissa Perrino, MD

Dana Tlais, MD

### Pathology

Jason Chiang, MD, PhD

David Ellison, MD, PhD

Julieann C. Lee, MD, MS

Brent A. Orr MD, PhD



**PBII**

Pediatric Biomedical  
Imaging Initiative

# 2025

## PEDIATRIC NEURO-ONCOLOGIC IMAGING EDUCATION COURSE

**October 30 - Nov 2, 2025**

St. Jude Children's Research Hospital  
Memphis, TN



Take advantage of early bird  
registration rates before August 31!  
**Scan QR code to register**

For more information, contact us:

[www.stjude.org/pbii-course](http://www.stjude.org/pbii-course)

[pbii@stjude.org](mailto:pbii@stjude.org)

Immerse yourself in a comprehensive 4-day course dedicated to the fundamentals and latest developments in pediatric neuro-oncology and neuro-oncologic imaging. Explore everything from neuropathology & advanced imaging to the most current management strategies aligned with the WHO 2021 CNS Tumor Classification.



# AJNR

## AMERICAN JOURNAL OF NEURORADIOLOGY

JULY 2025  
VOLUME 46  
NUMBER 7  
WWW.AJNR.ORG

Article Preview at [www.ajnr.org](http://www.ajnr.org) features articles released in advance.

### STATE OF PRACTICE

-   1289 **CPT Codes for MRI Safety—A User's Guide** *Colin M. Segovis, et al.*

HEALTH POLICIES/  
QUALITY  
IMPROVEMENT/  
EVIDENCE-BASED  
NEUROIMAGING

### REVIEW ARTICLES

-   1292 **A Review of the Opportunities and Challenges with Large Language Models in Radiology: The Road Ahead** *Neetu Soni, et al.*
-   1300 **Neuroimaging Spectrum of Erdheim-Chester Disease: An Image-Based Review** *Pranjal Rai, et al.*
-   1309 **Sodium MRI in Pediatric Brain Tumors** *Aashim Bhatia, et al.*

ARTIFICIAL INTELLIGENCE

BRAIN TUMOR IMAGING

PEDIATRIC NEUROIMAGING

### RADIOLOGY-PATHOLOGY CORRELATION

-   1318 **High-Resolution MR Imaging of the Parasellar Ligaments**  
*Ian T. Mark, et al.*

HEAD AND NECK IMAGING

### GENERAL CONTENTS

-   1321 **Cracking the Code of Calcification: How Presence and Burden among Intracranial Arteries Influence Stroke Incidence and Recurrence** *Matteo Conte, et al.*
-  1329 **Location-Specific Net Water Uptake and Malignant Cerebral Edema in Acute Anterior Circulation Occlusion Ischemic Stroke**  
*XiaoQing Cheng, et al.*
-   1336 **A Method for Imaging the Ischemic Penumbra with MRI Using Intravoxel Incoherent Motion** *Mira M. Liu, et al.*
-  1345 **Arterial Spin-Labeling MRI Identifies Abnormal Perfusion Metric at the Gray Matter/CSF Interface in Cerebral Small Vessel Disease**  
*Abdelkader Mahammedi, et al.*
- 1353 **Correlation between P2-PCA Volume Flow Rate and BOLD Cerebrovascular Reactivity in Patients with Symptomatic Carotid Artery Occlusion** *Amely Walser, et al.*

NEUROVASCULAR/  
STROKE IMAGING

-  1360 **Quantitative Shape Irregularity and Density Heterogeneity of Preoperative Hematoma Can Predict Rebleeding following Minimally Invasive Catheter Evacuation for Intracerebral Hemorrhage**  
*Kaijiang Kang, et al.* **NEUROVASCULAR/  
STROKE IMAGING**
-  1367 **Circle of Willis Variants and Their Association with Outcome after Successful Revascularization of Anterior Large Vessel Occlusion** *Aakanksha Sriwastwa, et al.* **NEUROVASCULAR/  
STROKE IMAGING**
-  1374 **Effect of SARS-CoV-2 Infection on Endovascular Thrombectomy Outcomes—Data from the Florida Stroke Registry**  
*Hayes B. Fountain, et al.* **NEUROINTERVENTION**
-  1379 **Hydrogel Coils versus Bare Platinum Coils for the Treatment of Ruptured and Unruptured Aneurysms: An Updated Systematic Review and Meta-Analysis of Randomized Controlled Trials**  
*Jonathan Cortese, et al.* **NEUROINTERVENTION**
-  1387 **Endovascular Treatment of Wide-Neck Intracranial Aneurysms Using the Novel Contour Neurovascular System: 5-Year Follow-up**  
*Howra Ktuyen, et al.* **NEUROINTERVENTION**
-  1395 **Safety and Effectiveness of a Novel Integrated Angiography System for Continuous Guidance and Flushing in Diagnostic Cerebral Catheter Angiography: A Randomized Controlled Trial** *Boseong Kwon, et al.* **NEUROINTERVENTION**
- 1403 **Reentry Technique for Rescue Recanalization of Carotid Near-Total Occlusion after Subintimal Penetration** *Dang Khoi Tran, et al.* **NEUROINTERVENTION**
-  1404 **Improving the Robustness of Deep Learning Models in Predicting Hematoma Expansion from Admission Head CT** *Anh T. Tran, et al.* **ARTIFICIAL INTELLIGENCE**
-  1412 **Deep Learning MRI Models for the Differential Diagnosis of Tumefactive Demyelination versus *IDH* Wild-Type Glioblastoma**  
*Gian Marco Conte, et al.* **ARTIFICIAL INTELLIGENCE**
-   1421 **Image-Based Search in Radiology: Identification of Brain Tumor Subtypes within Databases Using MRI-Based Radiomic Features**  
*Marc von Reppert, et al.* **BRAIN TUMOR IMAGING**
-  1429 **The Value of Quantitative Susceptibility Mapping and Morphometry in the Differential Diagnosis of Parkinsonism** *Yi Li, et al.* **NEURODEGENERATIVE  
DISORDER IMAGING**
-   1439 **Anti-amyloid Therapy and Cerebral Blood Flow Changes on MRI: A Potential Longitudinal Biomarker of Treatment Response?**  
*Andres Ricaurte-Fajardo, et al.* **NEURODEGENERATIVE  
DISORDER IMAGING**
-   1446 **Deep Invasion Volume of the Primary Nasopharyngeal Carcinoma Predicts Treatment Outcome** *Qi Yong H. Ai, et al.* **HEAD AND NECK IMAGING**
- 1453 **Peritumoral Hyperintense Signal on Postcontrast FLAIR Images Surrounding Vestibular Schwannomas following Stereotactic Radiosurgery** *Sandy T. Nguyen, et al.* **HEAD AND NECK IMAGING**
-  1458 **The Sensitivity of Arterial Spin-Labeling Imaging for Detection of Head and Neck Paragangliomas** *Yannan Yu, et al.* **HEAD AND NECK IMAGING**
-  1464 **Perfusion Showdown: Comparison of Multiple MRI Perfusion Techniques in the Grading of Pediatric Brain Tumors** *Chang Y. Ho, et al.* **PEDIATRIC NEUROIMAGING  
BRAIN TUMOR IMAGING**
-   1471 **Comparing CT-Like Bone Images Based on Fast Field Echo Resembling a CT Using Restricted Echo Spacing (FRACTURE) MR with CT in Pediatric Congenital Vertebral Anomalies** *Hirva Manek, et al.* **PEDIATRIC NEUROIMAGING**

 1478 Management Outcomes after Image-Guided Percutaneous Biopsy for Suspected Vertebral Osteomyelitis-Discitis *Dania G. Malik, et al.*

SPINE IMAGING AND  
SPINE IMAGE-GUIDED  
INTERVENTIONS

 1486 Prevalence and Characteristics of Microspurs in Patients with Spontaneous Intracranial Hypotension Compared with the General Population *Danial Nasiri, et al.*

SPINE IMAGING AND  
SPINE IMAGE-GUIDED  
INTERVENTIONS

 1493 Fusion of FDG and FMZ PET Reduces False-Positives in Predicting Epileptogenic Zone *Bingyang Cai, et al.*

MOLECULAR  
NEUROIMAGING/  
NUCLEAR MEDICINE

 1501 MRI Interpretation Errors in Adult Patients with Medically Refractory Epilepsy *Aoife M. Haughey, et al.*

ULTRA-HIGH-FIELD  
MRI/IMAGING OF  
EPILEPSY/DEMYELINATING  
DISEASES/INFLAMMATION/  
INFECTION

 1510 Cross-Sectional Validation of an Automated Lesion Segmentation Software in Multiple Sclerosis: Comparison with Radiologist Assessments *Maria Vittoria Spampinato, et al.*

ULTRA-HIGH-FIELD  
MRI/IMAGING OF  
EPILEPSY/DEMYELINATING  
DISEASES/INFLAMMATION/  
INFECTION

 1517 Application of Deep Learning Accelerated Image Reconstruction in T2-Weighted Turbo Spin-Echo Imaging of the Brain at 7T *Zeyu Liu, et al.*

ULTRA-HIGH-FIELD  
MRI/IMAGING OF  
EPILEPSY/DEMYELINATING  
DISEASES/INFLAMMATION/  
INFECTION



MR bone imaging technique, fast field echo resembling a CT using restricted echo spacing (FRACTURE), gives complementary information to conventional MRI of vertebral anomalies in children by providing osseous details, reducing the need for additional study, and avoiding radiation. A, Coronal CS FRACTURE MR MIP and (B) CT bone window images of a 5-year-old with KF anomaly show right occipital condyle hypoplasia (*white arrows*) and partial fusion of C2–C3 vertebrae (*black arrows*). C, Sagittal CS-FRACTURE MR and (D) CT bone window in a 6-year-old with KF anomaly show fusion of multiple cervical vertebrae (*blue arrows*), atlanto-occipital assimilation, hypoplastic dens, and basilar invagination (*black circle*). Acute angulation and effacement of the anterior CSF space at the cervicomedullary junction caused by the craniovertebral anomaly.

 Indicates Editor's Choices selection

 Indicates Fellows' Journal Club selection

 Indicates open access to non-subscribers at [www.ajnr.org](http://www.ajnr.org)

 Indicates article with supplemental online data

 Indicates article with supplemental online video

 Evidence-Based Medicine Level 1

 Evidence-Based Medicine Level 2

# CPT Codes for MRI Safety—A User's Guide

Colin M. Segovis, Jacob W. Ormsby, Cindy X. Yuan, Matthew J. Goette,  Melissa M. Chen, and  Heidi A. Edmonson



## ABSTRACT

**SUMMARY:** The magnetic fields of the MR environment present unique safety challenges. Medical implants and retained foreign bodies can prevent patients from undergoing MR imaging due to interactions between the magnetic fields of the MR environment and the implant or foreign body. These hazards can be addressed through careful MR safety screening and MR examination customization, often allowing these patients with implants to undergo management-altering MR imaging. However, mitigating these risks takes additional time, expertise, and effort. Effective in 2025, this additional work is formally acknowledged with a new series of Current Procedural Terminology codes to report the work of assessing and addressing safety concerns associated with implants and foreign bodies in the MR environment. This user guide provides guidance on how to report these codes so physician-led MR safety teams can be appropriately reimbursed for the additional work performed in preparing patients with implants or foreign bodies for MR imaging.

**ABBREVIATIONS:** ACR = American College of Radiology; ASTM = American Society for Testing Materials; CPT = Current Procedural Terminology; QHP = qualified health care professional

**M**RI is the mainstay technique of neuroimaging. The MR environment, however, presents unique safety challenges because of high-strength magnetic fields. Further, patients with implants can be at increased risk for injury in the MR environment. The US Food and Drug Administration (FDA) recognizes the risks of classifying medical implants as MR safe, MR conditional, or MR unsafe, as designated by the American Society for Testing Materials (ASTM) International, ASTM F2503-23 Standard Practice for Marking Medical Devices and Other Items for Safety in the MR Environment.<sup>1-3</sup> The MR safety of implanted medical devices is established by the manufacturer, with validation by the FDA. By default, any device that has not been deemed MR conditional or MR safe is considered MR unsafe. To further complicate matters, the term “MR nonconditional” has also been used in the medical literature to describe cardiac devices that have been labeled “MR unsafe” but may not be completely unsafe in the MR

environment.<sup>4,5</sup> The labeling provided by the manufacturer is not an absolute contraindication to obtaining an MRI with the guidance of a team knowledgeable in MR safety. Additionally, the number of medical implant technologies and market penetration of active medical implants continues to increase. MR safety is evolving as the field gains experience with implants and foreign bodies. However, the increasing complexity of MR conditional devices along with the time required to assess patient implants, modify medical device settings to meet manufacturers' MR conditions, and potentially modify MRI protocols to create a safe experience can require more time and effort than associated with the typical MRI, thus creating a challenge for the institution to offer MRI to patients with complex implants or foreign bodies.

Given the evolution of MR safety, the American College of Radiology (ACR) led the Current Procedural Terminology (CPT) code change application, along with the support of American Society of Neuroradiology and other radiology societies, which resulted in the creation of CPT codes that will be available for use in 2025.<sup>6,7</sup> These codes allow for billing for the additional work required to perform complex safety assessments for patients at increased risk for harm in the MR environment due to the presence of a medical implant or foreign body. These codes allow for reporting the work required to modify imaging protocols or implants to safely perform MR examinations. These codes describe the work of technologists, medical physicists, physicians, and/or other qualified health care professionals (QHPs). This paper provides guidance on how to use these codes in a neuroradiology practice.

Received November 6, 2024; accepted after revision January 11, 2025.

From the Department of Radiology and Imaging Sciences (C.M.S.), Emory University School of Medicine, Atlanta, Georgia; Department of Neuroradiology (J.W.O., M.M.C.), University of Texas MD Anderson Cancer Center, Houston, Texas; Department of Radiology (C.X.Y.), Indiana University School of Medicine, Indianapolis, Indiana; Enterprise Radiology (M.J.G.), Emory Healthcare, Atlanta, Georgia; and Department of Radiology (H.A.E.), Mayo Clinic, Rochester, Minnesota.

Please address correspondence to Colin M. Segovis, Emory Johns Creek Hospital, 6325 Hospital Pkwy, Johns Creek, GA 30097; e-mail: colin.m.segovis@emory.edu

 Indicates open access to non-subscribers at [www.ajnr.org](http://www.ajnr.org)

 Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8661>

## MR SAFETY CPT CODES

The MR safety CPT codes are designed to capture the work associated with specific tasks performed while planning for or performing an MR examination for a patient with an implant or foreign body, which is at high risk of interaction with the MR environment. Specific personnel perform and report these tasks, as described in Supplemental Data. CPT codes capture the work of MR technologists and medical physicists as clinical staff time, which is part of the technical component, while the physician's work is assigned to the professional component.<sup>7</sup> These codes are not meant to report the standard safety screening performed for all patients undergoing MR imaging, which is already included in each MR examination's reimbursement. Similarly, these codes should not be used for all patients with implants or foreign bodies since not all patients will require the additional work. The MR safety CPT codes should only be used when a safety situation requires extra work by the MR safety team outside of the facility's routine MR safety workflow, which includes MR conditional implants that may require precautions beyond normal operating mode of the MR system, implants lacking MR conditional labeling, or implants or foreign bodies that are typically contraindicated for MR.

The first 2 CPT codes in this family, 76014 and 76015, are used to report the MR safety assessment associated with an implant or foreign body, typically performed by an MR technologist. These 2 CPT codes are time-based and do not include any physician work. Code 76014 captures the first 15 minutes of work associated with the implant or foreign body assessment. As with all time-based CPT codes, this code can be reported when the midpoint (8 minutes) has passed. Code 76015 is used to report additional time after the initial 15 minutes, but the unit of time changes to 30 minutes. A minimum of 16 additional minutes of work is needed to report 76015 after the first 15 minutes captured with 76014. Code 76015 should not be reported without 76014, and 76015 can be reported up to 3 times for a given case as appropriate. Code 76015 would be expected to be used in scenarios requiring prolonged assessment time, such as with complex, multiple, or incompletely documented implants. These codes are typically used before the date of the MR examination; however, they can be used on the same day in urgent clinical situations. The MR technologist must document the relevant details of any implants and/or foreign bodies, along with the work/time in the medical record to report the codes.

CPT code 76016 is used to report an MR safety determination performed by a radiologist or other MR safety-trained QHP, typically performed before the day of the MRI examination. This code is expected to be reported infrequently, as it is only for clinical scenarios that require a risk-benefit analysis due to unclear MR conditions or when the examination may be contraindicated or limited by the presence of the implant or foreign body but important clinical information could be obtained from an MRI. Code 76016 is most likely performed by a radiologist but is not limited to radiologists. This code can be used on the same day as an MRI examination in urgent clinical scenarios. The code requires written documentation of the clinical scenario, risk-benefit analysis, potential alternate diagnostic tests, and recommendations to mitigate risk should the

MR examination remain the chosen diagnostic test. A possible result of the work of 76016 is that the ordered MR examination is not recommended.

The safe performance of MRI in the presence of implants, foreign bodies, or other safety concerns may require work on the day of the MR examination in addition to the work typically associated with MR imaging. Codes 76017, 76018, and 76019 are designed for the additional work associated with these safety activities.

Code 76017 captures the work of customizing the MR protocol to address safety concerns identified during the MR safety assessment. The work is performed by a medical physicist and/or MR safety expert to comply with implant-related MR requirements while maintaining/optimizing diagnostic image quality. This code is not time-based. The work is done in collaboration with the supervising physician or QHP. The work must be documented in the medical record by the physician or QHP; best practice may be to have an independent physics report or cosigned report (by a medical physicist and physician).

Code 76018 is used to report work done by the MR team on the same day as the MRI to prepare implanted electronics for the MRI. Many devices require changing the device to "MRI mode" or programming specific settings per the manufacturer's instructions. This code is not time-based. This code is separate from a cardiac device interrogation performed by cardiology or a neurostimulation analysis-programming service performed by neurology or neurosurgery and should be reported if the MR staff performs separate device programming. Modification of the settings of the device to minimize interactions between the device and MR environment is performed under the supervision of a qualified physician or QHP supervising the MR examination. Documentation of device modification by the physician or QHP overseeing the MR scan is required.

Code 76019 is used to report the work of positioning and/or immobilizing an implant before MR imaging. This can include the removal of a portion of the device or the immobilization of the device with a physical restraint. An example is the compression-wrapping of cochlear implants before an MRI per the manufacturer's instructions. Immobilization or positioning of the implant is done on the day of the examination and performed under the supervision of a physician or other QHP. It is best practice to follow the manufacturer's recommendations, including who the manufacturer suggests should position or immobilize the device. This code is not time-based. Documentation of the work is required by the supervising physician or QHP.

### Who Can Use the MR Safety Codes?

The MR safety codes should be used by individuals performing the work described by the codes, which can include MR technologists, MR safety experts, medical physicists, and physicians or QHPs. The personnel of the MR safety team are not defined by CPT. It is best practice to follow guidelines established by expert bodies such as the ACR Manual on MR Safety.<sup>2</sup> At a minimum, it is best practice that any individual performing an MR safety procedure is designated as "Level 2 Personnel" for the MRI facility.<sup>2</sup> Local and state regulations determine technologists' qualifications for the operation of an MRI. Supervision rules by a physician or

## MRI safety scenarios and associated CPT codes described in supplemental material

Scenario	Description	Associated Codes
1	MR safety evaluation performed by a technologist in advance of the examination	76014
2	MR safety evaluation performed by a technologist in advance of the examination	76014, 76015
3	MR safety evaluation performed by a technologist and radiologist	76014, 76016
4	MR safety evaluation performed by a technologist, physicist, radiologist in advance of the examination with examination customization on the day of the examination	76014, 76015, 76016, 76017
5	MR safety evaluation performed by a technologist in advance of the examination and in preparation of implant by the MR team on the day of the examination	76014, 76018
6	MR safety evaluation performed by a technologist in advance of the examination and immobilization of an implant by a qualified provider	76014, 76015, 76019
7	MR safety evaluation performed by a technologist and radiologist with a "no scan" recommendation	76014, 76016
8	MR safety evaluation performed by a technologist in advance of the examination	No code applicable

QHP are defined by Centers for Medicare & Medicaid Services (CMS). It is best practice for technologists to be registered as an MR technologist with a recognized credentialing body, such as the American Registry of Radiologic Technologists. Individuals may obtain board certification in MR safety, such as that offered by the American Board of Magnetic Resonance Safety. Board certification as a Magnetic Resonance Safety Officer, Magnetic Resonance Medical Director, or Magnetic Resonance Safety Expert is a best practice, but not required, to use the MR safety CPT codes.

### When to Use the MR Safety Code?

All patients must undergo a safety evaluation before MRI because of potential interactions between the high-strength magnetic fields of the MR environment and the patient. The MRI safety CPT codes are for situations when the evaluation associated with a given MRI scan exceeds the typical safety activities associated with MR imaging and should *not* be used as part of routine MR imaging. The MRI safety CPT codes are applicable to all clinical MR imaging regardless of field strength (eg, 1.5T, 3T, 7T). Example scenarios and coding guidance are meant to be general and are available in the Supplemental Data (Table). Individual practice patterns may vary.

### CONCLUSIONS

The MR safety CPT codes available starting in 2025 represent a recognition of the additional work required to safely perform MR examinations in patients with implants or foreign bodies. Radiologists, physicians, and other QHPs, medical physicists, and MR technologists engaged in MR safety activities will be able to report the extra effort/time needed to keep these patients safe. This paper provides guidance on the use of these new codes from the perspective of neuroradiologists, subject matter experts, and medical physicists engaged in MR safety and reimbursement. The use of these codes should be reserved for the preparation of patients with implants or foreign bodies at high risk for interaction with MR magnetic fields that require additional work

beyond the typical MR screening process. These codes are not inherently limited to radiology and can be used by individuals tasked with ensuring patient safety in the MR environment. Documentation in the medical record is required when reporting these codes. Documentation can be a distinct note or part of the diagnostic imaging report, depending on the facility's reporting workflow. MR imaging should be made available to as many patients as possible. These codes recognize the importance of access to MRI and the additional work required to ensure patients can safely undergo MR imaging.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

### REFERENCES

- American Society for Testing and Materials (ASTM). Designation: F2503-20). **Standard practice for marking medical devices and other items for safety in the magnetic resonance environment.** In: *Annual Book of ASTM Standards*. ASTM International; 2023
- ACR Committee on MR Safety. *ACR Manual on MR Safety*. American College of Radiology; 2024
- International Organization for Standardization. *Assessment of the Safety of Magnetic Resonance Imaging for Patients With an Active Implantable Medical Device*. 2018
- Indik JH, Gimbel JR, Abe H, et al. **2017 HRS expert consensus statement on magnetic resonance imaging and radiation exposure in patients with cardiovascular implantable electronic devices.** *Heart Rhythm* 2017;14: e97–153 CrossRef Medline
- Cunqueiro A, Lipton ML, Dym RJ, et al. **Performing MRI on patients with MRI-conditional and non-conditional cardiac implantable electronic devices: an update for radiologists.** *Clin Radiology* 2019;74: 912–17 CrossRef Medline
- American Medical Association. *CPT® 2025 Professional Edition*. American Medical Association; 2024
- United States, Health and Human Services, Centers for Medicare & Medicaid Services. **Medicare and Medicaid programs; CY 2025 payment policies under the physician fee schedule and other changes to part B payment and coverage policies; Medicare shared savings program requirements; Medicare prescription drug inflation rebate program; and Medicare overpayments.** 2024;42:405–91

# A Review of the Opportunities and Challenges with Large Language Models in Radiology: The Road Ahead

 Neetu Soni,  Manish Ora,  Amit Agarwal, Tianbao Yang, and  Girish Bathla



## ABSTRACT

**SUMMARY:** In recent years, generative artificial intelligence (AI), particularly large language models (LLMs) and their multimodal counterparts, multimodal large language models, including vision language models, have generated considerable interest in the global AI discourse. LLMs, or pre-trained language models (such as ChatGPT, Med-PaLM, LLaMA), are neural network architectures trained on extensive text data, excelling in language comprehension and generation. Multimodal LLMs, a subset of foundation models, are trained on multimodal data sets, integrating text with another modality, such as images, to learn universal representations akin to human cognition better. This versatility enables them to excel in tasks like chatbots, translation, and creative writing while facilitating knowledge sharing through transfer learning, federated learning, and synthetic data creation. Several of these models can have potentially appealing applications in the medical domain, including, but not limited to, enhancing patient care by processing patient data; summarizing reports and relevant literature; providing diagnostic, treatment, and follow-up recommendations; and ancillary tasks like coding and billing. As radiologists enter this promising but uncharted territory, it is imperative for them to be familiar with the basic terminology and processes of LLMs. Herein, we present an overview of the LLMs and their potential applications and challenges in the imaging domain.

**ABBREVIATIONS:** AI = artificial intelligence; BERT = bidirectional encoder representations from transformers; CLIP = contrastive language-image pre-training; FM = foundation models; GPT = generative pre-trained transformer; LLM = large language model; NLP = natural language processing; PLM = pre-trained language model; RAG = retrieval augmented generation; SAM = segment anything model; VLM = vision language model

The origin of language models dates from the 1990s with statistical language models focused on word prediction using n-grams and hidden Markov models. In 2013, neural language models like Word2Vec shifted the focus to distributed word embeddings using shallow neural networks.<sup>1</sup> The field, however, underwent a paradigm shift with the introduction of transformer architecture in 2017, based entirely on attention mechanism.<sup>2</sup> This was quickly followed by the introduction of pre-trained language models (PLMs), represented by bidirectional encoder representations from transformers (BERT) (2018) and BART (2019), which marked a major leap by utilizing transformers and context-aware word representations, greatly improving natural language processing (NLP) task performance.<sup>3,4</sup> More recently, it

was found that scaling PLM (in terms of model or data size), exemplified by generative pre-trained transformer (GPT)-3 (2020) and PaLM (2022), often leads to not only improved performance on downstream tasks but also some emergent abilities (eg, in-context learning and step-by-step reasoning) in solving a series of complex tasks. To differentiate these language models, the research community introduced the term “large language models” (LLMs) for the PLMs with massive size (eg, containing billions of parameters).<sup>5-7</sup>

LLMs have rapidly evolved since their introduction. These generative artificial intelligence (AI) models (including LLM, vision language models [VLMs], and diffusion-based models) can generate content in various domains, including language (GPT-4, PaLM, Claude), image (Midjourney, Stable Diffusion), codes (Copilot), and audio (VALL-E, resemble.ai).<sup>8-10</sup> Unlike traditional NLP models that process words sequentially, transformer-based models use attention layers to capture long-term dependencies. LLMs are trained on vast data and can produce human-like responses.<sup>11</sup> Publicly accessible ChatGPT was initially launched in 2022 by OpenAI<sup>12</sup>, followed by other LLMs such as Gemini, MedPaLM (Google), LLaVa-Med (Microsoft), Llama (Meta), and Claude 3 (Anthropic). These vary in the training parameters and purposes. While ChatGPT is a general-purpose LLM, MedPaLM and LLaVa-Med, for example, are tailored for medical applications. These models hold promise for enhancing radiology workflows by accelerating information retrieval,

Received September 26, 2024; accepted after revision November 13.

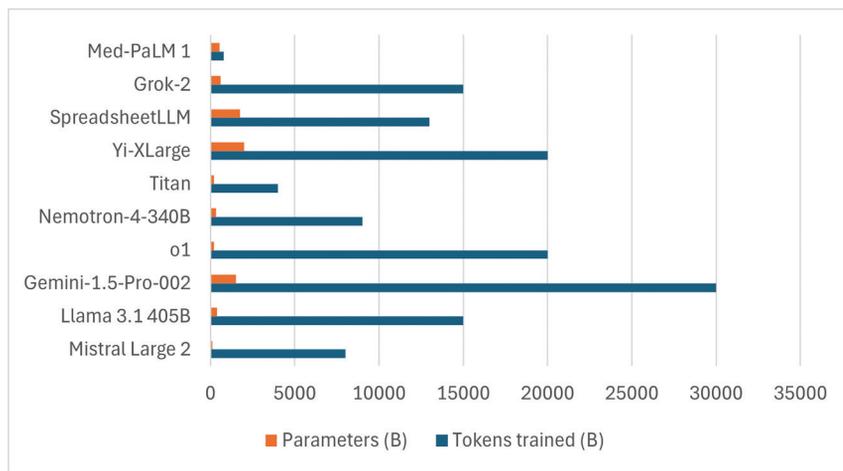
From the Department of Radiology (N.S., A.A.), Mayo Clinic—Jacksonville, Jacksonville, Florida; Department of Nuclear Medicine (M.O.), Sanjay Gandhi Post Graduate Institute of Medical Science, Lucknow, India; Department of Computer Science & Engineering (T.Y.), Texas A&M University, College Station, Texas; Department of Radiology (G.B.), Mayo Clinic, Rochester, Minnesota.

Please address correspondence to Neetu Soni, MD, FRCR, Department of Radiology, Mayo Clinic, 4500 San Pablo Rd, Jacksonville, FL 55902; e-mail: Soni.neetu@mayo.edu; @NeetuSo27437480

 Indicates open access to non-subscribers at [www.ajnr.org](http://www.ajnr.org)

 Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8589>



**FIG 1.** Bar chart showing the training tokens and parameters (in billions) of some of the common LLMs. Please note that several models are essentially part of larger families of models, and individual models may have variability in training tokens and parameters (data source: <https://lifecycle.ai/>).

generating comprehensive reports, and potentially aiding in diagnostic decision-making.<sup>12-14</sup> Given that radiology is often at the leading edge of technology and is associated with more than 70% of FDA-approved AI-enabled tools in the medical domain, it is unlikely that radiologists will remain untouched by this disruptive technology.<sup>15</sup>

Despite all the excitement around them, current LLMs are still in their infancy. Even though LLMs can mimic human conversations, they rely on word associations rather than true comprehension, limiting their problem-solving and logical reasoning abilities. LLMs may hallucinate or fabricate facts.<sup>16</sup> The successful integration of LLMs in radiology demands addressing critical challenges. They need vast training data sets and can give inconsistent responses because of their probabilistic nature. Addressing these challenges is paramount to fully realizing the potential of LLMs in radiology, which can be truly transformative. Herein, we briefly review the evolution of LLMs and their potential applications in radiology, including limitations, challenges, and possible solutions.

Prior to proceeding further, the interested reader is referred to the Supplemental Data for a glossary of common LLM-related terminology.

### MODELS, MODALS, AND MISCELLANEOUS THINGS IN BETWEEN!

A foundation model (FM) is an AI model trained using self-supervised learning with large unannotated data sets.<sup>17</sup> This confers broad capabilities to the model, enabling it to serve as a base (or foundation) for subsequent models. FMs trained on text, code, and images can be quite versatile. For example, the original ChatGPT (an LLM) was built on GPT 3.5 (a foundation model) and tweaked with chat-specific data.

LLMs are a subset of FMs that specialize in language tasks.<sup>18</sup> This is done by converting text into “tokens,” which are the fundamental units of input and output data. Tokenization is essential to understanding syntax, semantics, and relationships within the text, affecting eventual model performance and the ability to predict subsequent tokens.<sup>19,20</sup> Parameters, on the other hand, are trainable weights and biases within the model, which are learned

from the training data and can be considered the building blocks of the model. Figure 1 shows the number of training models and parameters of some of the common LLMs.

Even though LLMs work well for understanding and generating text, these are essentially unimodal, which limits their generalizability. However, integrating image identification/classification is challenging since most deep learning-based computer vision systems are data intensive and broadly not generalizable. While language is discrete and can be tokenized, visual concepts can evolve into higher dimensional spaces and can be difficult to discretize.<sup>21</sup> Naively discretizing images on a pixel-by-pixel basis may lose local neighbor-

hood information and may lead to prohibitive computing costs. For example, if the resolution of a color image is 256\*256\*3, the length of pixel tokens is 196,608. The use of self-attention size (a technique that identifies and weighs the various parts of an input sequence) can scale as a quadratic function of the token length and be computationally prohibitive. This prompted the development of VLMs, which can be broadly defined as multimodal models capable of performing inference with both images and text. The input may be image or text, while the output can be text, bounding box, or even segmentation masks. As of this writing, there are more than 112 publicly available open-source or application programming interface VLMs, including GPT-4v, Gemini, LLaVA, and others.<sup>22</sup> In general, VLMs are trained using four main strategies, either alone or in combination. In contrastive training, pairs of positive and negative examples are used, with the model trained to predict similar representations for the positive pairs. A typical example of this is contrastive learning image pre-training (CLIP), a neural network introduced in 2021.<sup>23</sup> CLIP combines a text and an image encoder and leverages information from two modalities to predict which caption goes with which image. The CLIP model has since been used for tasks such as generating images from text (Dalle-3, Midjourney), image segmentation tasks (Segment anything model [SAM]), and tasks involving image captioning and search. Note that CLIP is essentially adept at visual classification tasks. When provided with an image-text pair, it can determine if the two are a good fit for each other. However, it may not work well when differentiating categories with significant overlap.<sup>21</sup>

The second training strategy for VLMs is masking, where the VLM is trained to reconstruct missing patches in text (given an unmasked image) or vice-versa. The generative training paradigm, on the other hand, is used for models capable of generating entire images or very long captions, although some models may be trained to only generate images from text (eg, stable diffusion).<sup>24</sup> These models are generally more expensive to train. Finally, models using pre-trained backbones leverage open-source LLMs to learn the mapping between the image encoder and LLM (eg, Frozen, MiniGPT).<sup>21</sup> VLMs may also be further

subcategorized into models designed specifically for image interpretation and comprehension in conjunction with language (eg, CLIP), models that generate text from multimodal input (eg, GPT-4V), or models that can have both multimodal input and output (eg, Google Gemini). For a more detailed description of VLM, the interested reader is referred to the recent work by Ghosh et al.<sup>8</sup>

LLMs may perform the designated task after exposure to a few examples (few-shot learning), single examples (single-shot learning), or even without any training examples (zero-shot learning).<sup>5</sup> Another commonly used term in the field is “grounding,” which, in the context of LLMs, essentially implies providing the LLM with relevant and use case-specific information in order to obtain more accurate and relevant output. This is primarily done through retrieval augmented generation (RAG), which retrieves information (through databases, files, etc) and presents it to the LLM, along with the prompt. The LLM then uses this information while responding to the query.<sup>25</sup>

LLM models may be open-source or proprietary. Open-source LLMs, such as LLaMA series of models, have gained considerable popularity and attention in both academia and industry as a result of the available model checkpoints to customize, transparent model architecture, training process, data sets, and code. In contrast, closed-source LLMs, such as the ChatGPT family, only offer an application programming interface for users to access the LLMs instead of directly using the model. In particular, closed source options may also provide interfaces for users to further fine-tune released models on the host server. In evaluation, although closed-source models often tend to be more powerful because of their access to vast proprietary training data sets and advanced research resources, open-source models are still competitive with top-tier closed-source models.

LLMs, by nature, are probabilistic models, and the response can vary, even to the same query. This, in technical terms, is determined by the “temperature,” which can be adjusted based on the requirements. Models with a higher temperature give a more varied response, which can be entertaining but not ideal in the medical domain. Models with a temperature of zero are deterministic and always give essentially the same response to the same query.<sup>26</sup> Some authors have recently proposed using a context-aware temperature network, which can variably drive the temperature up or down, based on the context, eg, TempNet.<sup>27</sup> A summary of various LLMs used in the healthcare space is provided in Supplementary Data.

## POTENTIAL APPLICATIONS OF LLMs IN RADIOLOGY

LLMs have the potential to impact several facets of radiology, starting from study ordering and protocoling all the way to report generation and follow up. These can impact not only the radiologist but also the patient, primary healthcare providers, and the healthcare system. In the following sections, we briefly outline some of the potential LLM applications and associated challenges.

### Workflow Optimization

LLMs, given their ability to comprehend vast textual information (diagnostic requests, EHRs, prior imaging reports, guidelines, and medical literature)<sup>9,28</sup> can help with study protocoling in

routine and challenging cases.<sup>29</sup> Although tools like ChatGPT and Glass AI have shown promising results in this regard, a study by Nazario-Johnson et al,<sup>30</sup> noted that their accuracy currently lags behind that of experienced neuroradiologists.

Chatbots based on LLMs may also be used to provide education about CT/MRI procedures in common terms, thereby reducing patient anxiety while improving patient understanding and engagement.<sup>31</sup> Even though LLM responses are generally accurate, they are currently not perfect and require oversight. Also, GPT-4 has been utilized to create summaries and graphical representations of disease courses from the previous MRI reports in patients with glioblastoma, which can potentially save time when comparing multiple prior studies.<sup>32</sup>

### Image Segmentation

FMs can be helpful in reducing the burden associated with manual segmentations, which are labor-intensive and require significant expertise. Unlike deep learning-based semi- or fully automatic segmentation methods, which can have limited generalizability, foundation segmentation models are more broadly generalizable.<sup>33</sup> SAM, a segmentation model with zero-shot generalization, generates masks for objects in natural images with distinct boundaries.<sup>34,35</sup> Combining SAM with localization algorithms or integrating it with image processing tools like 3D Slicer enhances its medical imaging applications.<sup>36</sup> MedSAM is trained on over 1 million medical image-mask pairs from 10 imaging modalities and more than 30 cancer types. MedSAM demonstrates accurate segmentation, achieving results comparable with or better than models like U-Net and DeepLabV3+.<sup>37</sup> It could be used for 3D tumor annotation and assessing treatment responses.<sup>38</sup> A more recent update, SAM2 is capable of not only segmenting 2D images, but also 3D data sets and videos.<sup>39</sup> A more recent addition to the list of segmentation models is CT Foundation, a CT-based model developed for 3D segmentation across different body parts. The model was launched recently by Google and was trained using over a half-million de-identified CT volumes.<sup>40</sup> These models could serve as a one-stop shop in the future for radiology-specific tasks instead of having multiple separate segmentation models for individual pathologies (such as glioma, meningioma, and vestibular schwannoma).

### Image Interpretation and Report Generation

Some prior studies have noted a superior diagnostic performance of Claude 3 Opus over GPT-4o and Gemini 1.5 Pro in “Diagnosis Please” radiology cases.<sup>41</sup> Similarly, GPT-4 Turbo was used to analyze 751 neuroradiology cases from the *American Journal of Neuroradiology* with an initial diagnostic accuracy of 55.1%, which improved to 72.9% with customized prompt engineering.<sup>42</sup> Another recent study compared ChatGPT-4V and Gemini Pro Vision with radiologists and noted 49%, 39%, and 61% accuracy across 190 radiology cases, respectively.<sup>43</sup> Finally, models like Bard, ChatGPT-3.5, and GPT-4 have outperformed human consensus and MedAlpaca by at least 5% and 13%, respectively, for rare and complex diagnoses, with GPT-4 achieving a diagnostic accuracy of 93%.<sup>44</sup> However, an important caveat here is that all these studies used either history and/or curated limited images

per case, as is often the case with online educational content. Unless the radiologist hand-picks individual images for the model to evaluate, along with prompt engineering, the large-scale automated generalization capability of these models is unclear. Similarly, Liu et al<sup>45</sup> recently proposed a novel framework for generating radiology reports from high-resolution 3D CT chest data sets without image down-sampling, again showing the potential application of VLMs in 3D-data sets. More simplified LLMs for image interpretation and report generation have also been proposed for 2D images such as chest radiographs.<sup>45,46</sup> None of these 2D and 3D models, however, have been extensively evaluated prospectively to ensure fairness, lack of bias, or ability to detect rare diseases, which are important considerations for future validations.

Radiology reports are often written in a freestyle format, which may hinder the extraction of meaningful information for clinical or research purposes.<sup>47</sup> LLMs have also been deployed to generate radiology reports by structuring sections such as findings, impressions, and differential diagnoses while integrating demographic data and keywords.<sup>48</sup> In a recent study, GPT-4 showed excellent accuracy in selecting the most appropriate report template and identifying critical findings.<sup>49</sup> Another study noted AI-structured reports to be comparable to those generated by radiologists, often outperforming the latter in clarity, brevity, and ease of understanding.<sup>50</sup>

Another retrospective study compared the detection of common reporting errors (eg, omission, insertion, spelling mistakes) by testing a curated data set of erroneous reports. GPT-4 demonstrated a detection rate similar to senior radiologists, attending physicians, and residents (detection rate, 82.7%, 89.3%, 80.0%, and 80.0%, respectively), albeit with reduced time and cost.<sup>51</sup> LLMs have also shown promise in automated TNM classification in lung cancer staging based solely on radiology reports without additional training, especially when provided with TNM definitions.<sup>52</sup> Finally, LLMs have also shown promise in terms of coding radiology reports and suggesting follow-up based on the presence or absence of pathologies (eg, aortic aneurysm) or employing coding systems (such as Lung-RADS) to screening CT studies.<sup>53</sup>

### **Patient-Oriented Reports**

LLMs have been shown to simplify radiology report impressions, making them more comprehensible to patients.<sup>54</sup> For example, a study comparing LLM-generated MRI spine reports with original reports found that the former had higher comprehension scores among both radiologists and non-physician raters.<sup>55</sup> However, these AI-generated reports still require edits and expert supervision, often to remove irrelevant suggestions of causality, prognosis, or treatment.<sup>56</sup> Models using patient-friendly language with illustrations of hyperlinked terms perform even better in terms of patient comprehension.<sup>55,56</sup> Similarly, some of the recently released LLMs (Med-Flamingo, LLaVA-Med) have shown promising results in visual question answering and rationale generation, which can augment responses to patient queries about reports, implications, and follow-up.<sup>57,58</sup>

### **Clinical/Tumor Board Decision-Making in the Future**

The role of ChatGPT has also been explored for glioma adjuvant therapy decisions in tumor boards. A small case study ( $n = 10$ )

noted that the LLM-provided recommendations were rated moderate to good by the experts, even though the model performed poorly in classifying glioma types and lacked sufficient precision to replace expert opinion.<sup>59</sup>

### **Radiology Training and Research**

LLMs can potentially help simplify complex scientific schematics, compare radiological images, and reduce repetitive tasks and activities that may help in radiology education.<sup>60</sup> Another study noted that LLMs like Vicuna-13B can identify various findings on chest radiography reports and show moderate to substantial agreement with existing labelers across data sets like MIMIC-CXR and NIH.<sup>26</sup> The LLMs can, therefore, help curate larger data sets while reducing human efforts.

LLMs have also been explored for reviewing manuscripts. The peer review feedback for GPT-4 has shown a considerable overlap (31%) with human reviewers (comparable to overlap between different human reviewers), with users rating it as more beneficial than human feedback.<sup>61</sup> LLMs have also been explored for text summarization and editing, especially for non-native authors.<sup>62</sup> Use of LLMs for manuscript writing is a big ethical concern and can be a threat not only to the credibility of the paper but also to the authors and the journal itself. Most journals currently do not allow the LLM to be designated as a co-author and ask for transparency from the authors in terms of declaring any use of the LLM in manuscript preparation. Note that LLMs are not databases and are designed to be used as a general reasoning and text engine. They are also well-known to hallucinate references. Some of these problems may be partially overcome with LLMs trained specifically for academic pursuits (eg, Scispace), which can allow the user to interact with pre-selected papers to understand complex research better, extract relevant information, and identify gaps in existing knowledge.<sup>63</sup> When used in an ethical way, these resources can potentially enhance the impact of a researcher's work. Such ethical use is not always a given, and these developments present a more challenging landscape to journals and editors.

### **LLM LIMITATIONS AND DRAWBACKS**

Despite the impressive performance of LLMs, there are several limitations and potential risks. A Delphi study highlighted concerns among researchers regarding cybersecurity breaches, misinformation, ethical dilemmas, biased decision-making, and inaccurate communication.<sup>19</sup> LLMs generate responses based on statistical pattern recognition, lacking a deep contextual understanding of medical concepts, which can result in errors.<sup>64,65</sup> They often fail at common sense reasoning, leading to incorrect or biased outputs. A recent work noted that LLMs can be rather fragile in mathematic reasoning and argued that the current LLMs may not be capable of genuine mathematical reasoning.<sup>66</sup> LLMs may produce plausible yet incorrect information "hallucinations" in diagnostics and report generation.<sup>67,68</sup>

Biases can also arise from failure to capture the complexity of real-world clinical scenarios. This can lead to significant inaccuracies, especially for rare diseases, under-represented groups, third-world populations, and non-English literature. LLMs may perpetuate biases from their training data, leading to misinterpretations

and inappropriate treatment recommendations.<sup>69</sup> Privacy concerns exist as LLMs are trained on large data sets that may include sensitive patient information. The risk of disclosing such information without consent is considerable.<sup>70,71</sup>

LLMs can generate convincing but misleading explanations for incorrect answers, known as “adversarial helpfulness.” This can deceive humans and other LLMs by reframing questions, showing overconfidence, and presenting selective evidence. This highlights the need for caution because of the opaque nature of LLMs, which complicates transparency and understanding of their decision-making processes.<sup>2,64,72,73</sup> LLMs may struggle to maintain context over long passages, leading to disjointed responses, and they might not be up-to-date with proprietary information or recent advancements.<sup>74</sup> LLMs are also vulnerable to adversarial attacks, where malicious inputs deceive the model into producing harmful outputs and reveal confidential information.<sup>75</sup>

In terms of more specific limitations pertaining to the medical domain, it is unlikely that a single foundational model can serve as a go-to resource given the number of known and continuous newly defined entities, various imaging modalities, and their own inherent resolutions, utilities, and limitations.<sup>76</sup> It is also unclear whether a domain-specific LLM versus a modality-specific LLM might be a better long-term solution. Another issue is the lack of high-quality annotations in the medical domain, especially 3D data sets, which limits the amount of available training data. Similarly, the inherent imbalance in the real-world data cannot be ignored. Rare diseases are, by definition, under-represented. Lack of sufficient training data can lead to performance degradation later. Given the dynamic nature of the medical domain, it is inevitable that such models would require continuous retraining and validation. However, occasionally, models may lose previously acquired capabilities while acquiring new ones, as happened with GPT-4 (March 2023), which could differentiate between prime and composite numbers with reasonable accuracy but showed poor performance on the same questions subsequently (GPT-4, June 2023).<sup>77</sup> Finally, even though RAG has been shown to alleviate some of the shortcomings of LLMs by grounding and providing context-specific information, the relative prevalence of redundant pieces of information may suppress more recent sparse yet critical information and lead to incorrect responses.<sup>78</sup> A hypothetical example would be a recent change in tumor classification or treatment strategy for a certain disease(s). Given the recent change, the information may not be prevalent in the literature and thus be ignored by RAG and LLM while formulating a response. In research, LLMs face limitations due to hallucinations, data bias, misinformation, and a lack of transparency. The training data often lag recent advancements, leading to outdated insights. Also, the resource costs and environmental impact of running LLMs are non-trivial. Over-reliance on these models may erode researchers’ critical thinking and problem-solving skills. Ethical concerns also arise regarding privacy, copyright, and plagiarism, as LLMs cannot be held accountable or listed as authors.<sup>79</sup> Consequently, journal guidelines now often require the disclosure of LLM use in manuscript preparation to ensure transparency and maintain the integrity of the review process.<sup>80</sup>

## FUTURE DIRECTIONS

The US Food and Drug Administration (FDA) has authorized about 1000 AI-enabled medical devices but has yet to authorize an LLM despite acknowledging their potential to positively impact healthcare. Given the complexity of LLMs and the possible output permutations, the FDA recognizes the need for regulatory innovation and specialized tools that allow LLM evaluation in the appropriate context and settings.<sup>81</sup> As noted with the various aspects of radiology, the LLM performance currently has considerable challenges in terms of addressing model reliability, explainability, accountability, consistently matching expert-level performance, and withstanding rigorous scrutiny. Even though the news of an LLM outperforming an expert on a test may be eye-catching, radiologists impact several facets of patient care simultaneously in a very dynamic field, and the role of trained medical professionals cannot be taken lightly.

Improving the explainability and generalizability of LLMs is essential for building human trust, given the current limitations.<sup>82</sup> Active involvement of domain experts in data selection and model fine-tuning ensures that LLM-generated insights are reviewed and validated before application in patient care, thereby improving accuracy and reducing errors.<sup>64</sup> Model training should also address the real-life challenges of imbalanced data and rare diseases, which may considerably impact eventual model performance. This may be done by privacy-compliant data sharing to mitigate real-world data scarcity, limiting the use of synthetic or augmented data (especially for rare cases), and ensuring overall high-quality ground-truth data. It is important to note here that over-reliance on synthetic data can be problematic because it often lacks the complexity of actual data and can lead to model collapse in a real-life setting.<sup>83</sup> Patient privacy concerns should be addressed from the training stage itself by excluding any patient-specific identifiers like name, address, and medical record numbers.

The LLM design for the medical domain should also consider the need for tighter regulatory compliances in this field. Models that can provide confidence scores, generating receiver-operating-characteristics curves, are explainable and trained to be fair in terms of patient gender, race, and age and are more likely to survive regulatory scrutiny. Ethical concerns must be addressed to prevent perpetuating biases from training data.<sup>69</sup> Compliance with regulations like the Health Insurance Portability and Accountability Act of 1996 (HIPAA) is essential for maintaining data privacy and regulatory compliance.<sup>84</sup> Open-source LLMs that can operate locally without sharing data with third parties are a promising privacy-preserving alternative.<sup>85</sup> Open-source models (eg, LLaMA) are less reliant on proprietary data sources, potentially increasing transparency and accessibility.<sup>14</sup> For proprietary models, additional considerations with regard to the source and quality of training data and any related copyright issues also need to be addressed prior to implementation.

Validating the information generated by LLMs is essential and requires regular audits, fairness-aware training, and ethical guidelines.<sup>64,67,86</sup> Paraphrasing a question or providing additional context to an LLM can change the subsequent response.<sup>83</sup> Hence, the validation needs to be not only on scientific rigor but also on the contextual understanding of LLM. For example, a rounded peripheral hyper density on a non contrast CT may reflect a

contusion (in the context of trauma), a metastatic lesion (in a patient with a known malignant melanoma), a spontaneous hemorrhage (in an older patient with amyloid angiopathy), or a hemorrhagic venous infarct (in a young female on oral contraceptives). Understanding the clinical context in such cases is critical. Traditional scores of accuracy and performance metrics, therefore, may not fully evaluate such models. Similarly, LLM performance may change over time (LLM drift) and can be especially troubling for proprietary models where little is known about the underlying architecture and training data used.<sup>83</sup> Additional validation should include testing the LLM on a mix of population cohorts to ensure the model performance is similar with regard to the patient's gender, demographics, and geographical distribution, ensuring model fairness and lack of bias. LLMs often struggle with outdated information because of static training data. Not only is it important for the LLM to be able to retrieve information from a continuously updated database, but it should also be able to prioritize more recent critical changes over redundant but over-represented literature. Combining LLMs with external data retrieval systems can enhance content generation.<sup>87</sup> For example, RadioRAG, a model developed to retrieve real-time information from online resources (like *Radiopaedia*), variably improved the performance of various LLMs.<sup>88</sup> The recent launch of OpenAI o1-preview series aims to tackle complex reasoning tasks by allowing the model to spend more time thinking before responding. This approach helps align it more closely with ethical guidelines and reduces the risk of unsafe or biased content. This model demonstrates expert-level performance on challenging problems by incorporating advanced reasoning capabilities to adhere better to safety protocols and ensure thoughtful, reliable outputs.<sup>89</sup>

One must also consider the more practical challenges to LLM implementation, including the need for additional energy and infrastructure, resources to ensure continued compliance with regulatory standards, potential safety risks, and medicolegal implications that may offset any efficiency gains. There is also a lack of clarity regarding the cost structure, as some companies may charge based on a number of tokens while others may charge based on hours of usage. Additional costs related to network usage, embedding (use of RAG), and periodic learning will also need to be considered.<sup>90</sup> It is also unclear if LLM behavior may change with software or scanner upgrades or the introduction of newer sequences. Finally, establishing mechanisms for fixing liability when an incorrect model decision impacts patient care is also critical and requires both local and national coordination to be uniformly implemented.

In terms of the role of LLMs in neuroradiology, given the uncharted territory, it may be helpful to first validate these models in lower-risk clinical workflows where the LLM output may be annoying or unhelpful but not detrimental to patient health. These may include study protocolling for common exams, summarizing medical history or prior MRI reports, lesion segmentation, and volumetry. Such outputs are open to validation and allow for a more nuanced model evaluation. Using LLMs for lesion characterization or differential diagnosis generation sounds exciting but can pose considerable challenges in real-life settings. It is also important that the model is interpretable. For this, the model would

provide not only the possible differential considerations but also the factors that the models considered and how the model weighed them. Another important consideration is rigorous model testing, as model performance may be impacted by model size, domain-specific nature, prompt engineering, and optimization.<sup>91</sup> By focusing on these areas in the near future, the field can make early inroads in developing AI solutions that effectively address the complex needs of radiology.

The current challenges in the LLM field also underscore the need for continued short- and long-term research into this field to ensure LLMs are fully utilized. These would include further work into LLMs that are fair, ethical, equitable, and unbiased. Mechanisms that improve model explainability, allow inherent safety guardrails, and minimize or stop hallucinations would further improve user trust. Similarly, further research is needed to find ways for an LLM to continuously update with relevant literature without necessarily forgetting prior information and explore new methods to identify model performance degradation. Equally importantly, further research into new and innovative methods of model validation that use a multi faceted approach beyond traditional performance metrics is needed.

At this time, it is difficult to predict the eventual extent and scale of disruption that LLMs may cause and how they might reshape the role of radiologists in the future. It is possible that LLMs may reduce or eliminate the need for mundane tasks such as study protocolling for common indications, make radiology reports more objective through volumetric inputs, reduce radiologist effort by summarizing impressions, or help with clinical workflows such as patient scheduling, summarizing patient history, and treatment details. LLMs may also play an important role in trainee education, simplifying complex topics, or in research by helping with data collection or annotation. LLMs, in essence, have vast unrealized potential that is dependent on how well the existing challenges are addressed.

## CONCLUSIONS

LLMs have transformative potential in radiology with several potential medical applications, but their effective implementation requires addressing key limitations. Researchers and healthcare professionals must navigate these limitations and employ innovative solutions to maximize LLM effectiveness. The future of LLMs in radiology lies in addressing these challenges through interdisciplinary collaboration, ongoing research, and the development of ethical, transparent, and privacy-compliant AI systems. Successful clinical implementation of LLMs would require considerable coordination between domain experts in medicine and computer sciences, researchers, and industry and regulatory authorities.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Mikolov T, Sutskever I, Chen K, et al. **Distributed representations of words and phrases and their compositionality.** *Adv Neural Inf Process Syst* 2013;26:3111–3119
2. Vaswani A, Shazeer N, Parmar N, et al. **Attention is all you need.** *Adv Neural Inf Process Syst* 2017;30:6000–10

3. Devlin J, Chang M-W, Lee K, et al. **BERT: pre-training of deep bidirectional transformers for language understanding.** In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* Association for Computational Linguistics; 2019:4171–86
4. Lewis M, Liu Y, Goyal N, et al. **BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension.** In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics; 2020:7871–80
5. Brown T, Mann B, Ryder N, et al. **Language models are few-shot learners.** *Adv Neural Inf Process Syst* 2020;33:1877–1901
6. Chowdhery A, Narang S, Devlin J, et al. **Palm: Scaling language modeling with pathways.** *J Mach Learn Res* 2023;24:11324–43
7. Zhao WX, Zhou K, Li J, et al. **A survey of large language models.** *arXiv preprint arXiv:230318223* 2023
8. Ghosh A, Acharya A, Saha S, et al. **Exploring the frontier of vision-language models: a survey of current methodologies and future directions.** *arXiv preprint arXiv:240407214* 2024
9. Akinci D'Antonoli T, Stanzione A, Bluethgen C, et al. **Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions.** *Diagn Interv Radiol* 2024;30:80–90 CrossRef Medline
10. Bhayana R, Krishna S, Bleakney RR. **Performance of ChatGPT on a radiology board-style examination: insights into current strengths and limitations.** *Radiology* 2023;307:e230582 CrossRef Medline
11. Liu Z, Zhong A, Li Y. **Radiology-GPT: a large language model for radiology.** *Meta-Radiology* Epub ahead of print May 16, 2025
12. Gallifant J, Fiske A, Levites Strelakova YA, et al. **Peer review of GPT-4 technical report and systems card.** *PLOS Digital Health* 2024;3:e0000417 CrossRef Medline
13. Anil R, Dai AM, Firat O, et al. **Palm 2 technical report.** *arXiv preprint arXiv:230510403* 2023
14. Touvron H, Lavril T, Izacard G, et al. **LLaMA: open and efficient foundation language models.** *arXiv preprint arXiv:230213971* 2023
15. Joshi G, Jain A, Araveeti SR, et al. **FDA-approved artificial intelligence and machine learning (AI/ML)-enabled medical devices: an updated landscape.** *Electronics* 2024;13:498 CrossRef
16. Banerjee S, Agarwal A, Singla S. **LLMs will always hallucinate, and we need to live with this.** *arXiv preprint arXiv:240905746* 2024
17. Wiggins WF, Tejani AS. **On the opportunities and risks of foundation models for natural language processing in radiology.** *Radiol Artif Intell* 2022;4:e220119 CrossRef Medline
18. Paaß G, Giesselbach S. *Foundation Models for Natural Language Processing: Pre-Trained Language Models Integrating Media.* Springer-Verlag Nature; 2023
19. Denecke K, May R, Rivera Romero O. **Potential of large language models in health care: Delhi study.** *J Med Internet Res* 2024;26:e52399 CrossRef Medline
20. Yang X, Chen A, PourNejatian N, et al. **Gatortron: a large clinical language model to unlock patient information from unstructured electronic health records.** *arXiv preprint arXiv:220303540* 2022
21. Bordes F, Pang RY, Ajay A, et al. **An introduction to vision-language modeling.** *arXiv preprint arXiv:240517247* 2024
22. OpenVLM Leaderboard. 2024
23. Radford A, Kim JW, Hallacy C, et al. **Learning transferable visual models from natural language supervision.** Proceedings of the 38th International Conference on Machine Learning. *Proc Mach Learn Res* 2021;139:8748–63
24. Rombach R, Blattmann A, Lorenz D, et al. **High-resolution image synthesis with latent diffusion models.** Paper presented at: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; June 21–24, 2022; New Orleans, LA
25. Berger E. **Grounding LLMs. Eleanor on Everything.** Accessed January 6, 2025. <https://everything.intellectronica.net/p/grounding-llms>
26. Mukherjee P, Hou B, Lanfredi RB, et al. **Feasibility of using the privacy-preserving large language model Vicuna for labeling radiology reports.** *Radiology* 2023;309:e231147 CrossRef Medline
27. Qiu Z-H, Guo S, Xu M, et al. **To cool or not to cool? Temperature network meets large foundation models via DRO.** *arXiv preprint arXiv:240404575* 2024
28. López-Úbeda P, Martín-Noguerol T, Juluru K, et al. **Natural language processing in radiology: update on clinical applications.** *J Am Coll Radiol* 2022;19:1271–85 CrossRef Medline
29. Gertz RJ, Bunck AC, Lennartz S, et al. **GPT-4 for automated determination of radiological study and protocol based on radiology request forms: a feasibility study.** *Radiology* 2023;307:e230877 CrossRef Medline
30. Nazario-Johnson L, Zaki HA, Tung GA. **Use of large language models to predict neuroimaging.** *J Am Coll Radiol* 2023;20:1004–09 CrossRef Medline
31. Kuckelman IJ, Paul HY, Bui M, et al. **Assessing AI-powered patient education: a case study in radiology.** *Acad Radiol* 2024;31:338–42 CrossRef Medline
32. Laukamp KR, Terzis RA, Werner JM, et al. **Monitoring patients with glioblastoma by using a large language model: accurate summarization of radiology reports with GPT-4.** *Radiology* 2024;312:e232640 CrossRef Medline
33. Antonelli M, Reinke A, Bakas S, et al. **The medical segmentation decathlon.** *Nat Commun* 2022;13:4128 CrossRef Medline
34. Kirillov A, Mintun E, Ravi N, et al. **Segment anything.** *Proceedings of the IEEE/CVF International Conference on Computer Vision*; October 1–6, 2023; Paris, France: 4015–26
35. Huang Y, Yang X, Liu L, et al. **Segment anything model for medical images?** *Med Image Anal* 2024;92:103061 CrossRef Medline
36. Liu Y, Zhang J, She Z, et al. **SAMM (segment any medical model): a 3D slicer integration to SAM.** *arXiv preprint arXiv:230405622* 2023
37. Ma J, He Y, Li F, et al. **Segment anything in medical images.** *Nat Comm* 2024;15:654 CrossRef Medline
38. Ma J, Wang B. **Towards foundation models of biological image segmentation.** *Nat Methods* 2023;20:953–55 CrossRef Medline
39. Ravi N, Gabeur V, Hu Y-T, et al. **SAM 2: segment anything in images and videos.** *arXiv preprint arXiv:240800714* 2024
40. Kiraly A, Traverse M; Google research. **Taking medical imaging embeddings 3D.** October 21, 2024. Accessed February 28, 2025. <https://research.google/blog/taking-medical-imaging-embeddings-3d/>
41. Sonoda Y, Kurokawa R, Nakamura Y, et al. **Diagnostic performances of GPT-4o, Claude 3 Opus, and Gemini 1.5 Pro in “diagnosis please” cases.** *Jpn J Radiol* 2024;42:1231–35 CrossRef Medline
42. Wada A, Akashi T, Shih G, et al. **Optimizing GPT-4 turbo diagnostic accuracy in neuroradiology through prompt engineering and confidence thresholds.** *Diagnostics (Basel)* 2024;14:1541 CrossRef Medline
43. Suh PS, Shim WH, Suh CH, et al. **Comparing diagnostic accuracy of radiologists versus GPT-4V and Gemini Pro Vision using image inputs from diagnosis please cases.** *Radiology* 2024;312:e240273 CrossRef Medline
44. Abdullahi T, Singh R, Eickhoff C. **Learning to make rare and complex diagnoses with generative AI assistance: qualitative study of popular large language models.** *JMIR Med Educ* 2024;10:e51391 CrossRef Medline
45. Liu C, Wan Z, Wang Y, et al. **Benchmarking and boosting radiology report generation for 3D high-resolution medical images.** *arXiv preprint arXiv:240607146* 2024
46. Wang Z, Liu L, Wang L, et al. **R2GenGPT: radiology report generation with frozen LLMs.** *Meta-Radiology* 2023;1:100033 CrossRef
47. Pinto Dos Santos D, Cuocolo R, Huisman M. **O structured reporting, where art thou?** *Eur Radiol* 2024;34:4193–94 CrossRef Medline
48. Hartung MP, Bickle IC, Gaillard F, et al. **How to create a great radiology report.** *Radiographics* 2020;40:1658–70 CrossRef Medline
49. Adams LC, Truhn D, Busch F, et al. **Leveraging GPT-4 for post hoc transformation of free-text radiology reports into structured reporting: a multilingual feasibility study.** *Radiology* 2023;307:e230725 CrossRef Medline

50. Hasani AM, Singh S, Zahergivar A, et al. **Evaluating the performance of Generative Pre-trained Transformer-4 (GPT-4) in standardizing radiology reports.** *Eur Radiol* 2024;34:3566–74 CrossRef Medline
51. Gertz RJ, Dratsch T, Bunck AC, et al. **Potential of GPT-4 for detecting errors in radiology reports: implications for reporting accuracy.** *Radiology* 2024;311:e232714 CrossRef Medline
52. Matsuo H, Nishio M, Matsunaga T, et al. **Exploring multilingual large language models for enhanced TNM classification of radiology report in lung cancer staging.** *arXiv preprint arXiv:240606591* 2024
53. Yan A, McAuley J, Lu X, et al. **RadBERT: adapting transformer-based language models to radiology.** *Radiol Artif Intell* 2022;4:e210258
54. Doshi R, Amin KS, Khosla P, et al. **Quantitative evaluation of large language models to streamline radiology report impressions: a multimodal retrospective analysis.** *Radiology* 2024;310:e231593 CrossRef Medline
55. Park J, Oh K, Han K, et al. **Patient-centered radiology reports with generative artificial intelligence: adding value to radiology reporting.** *Sci Rep* 2024;14:13218 CrossRef Medline
56. Berigan K, Short R, Reisman D, et al. **The impact of large language model-generated radiology report summaries on patient comprehension: a randomized controlled trial.** *J Am Coll Radiol* 2024;21:1898–903 CrossRef Medline
57. Li C, Wong C, Zhang S, et al. **LLaVA-Med: training a large language-and-vision assistant for biomedicine in one day.** In: *Proceedings of the 37th International Conference on Neural Information Processing Systems (NIPS '23)*. Curran Associates Inc; 2023;1240:28541–64
58. Moor M, Huang Q, Wu S, et al. **Med-flamingo: a multimodal medical few-shot learner.** *Proceedings of the 3rd Machine Learning for Health (ML4H)*. *Proc Mach Learn Res* 2023;225:353–67
59. Haemmerli J, Sveikata L, Nouri A, et al. **ChatGPT in glioma adjuvant therapy decision making: ready to assume the role of a doctor in the tumour board?** *BMJ Health Care Inform* 2023;30:e100775 CrossRef Medline
60. Tippareddy C, Jiang S, Bera K, et al. **Radiology reading room for the future: harnessing the power of large language models like ChatGPT.** *Curr Probl Diagn Radiol* 2023 August 30. [Epub ahead of print] CrossRef Medline
61. Liang W, Zhang Y, Cao H, et al. **Can large language models provide useful feedback on research papers? A large-scale empirical analysis.** *N Eng J Med AI* 2024;1:A10a2400196 CrossRef
62. Hwang SI, Lim JS, Lee RW, et al. **Is ChatGPT a “fire of Prometheus” for non-native English-speaking researchers in academic writing?** *Korean J Radiol* 2023;24:952–59 CrossRef Medline
63. Scispace. **Scispace. The Fastest Research Platform Ever**
64. Ullah E, Parwani A, Baig MM, et al. **Challenges and barriers of using large language models (LLM) such as ChatGPT for diagnostic medicine with a focus on digital pathology—a recent scoping review.** *Diagn Pathol* 2024;19:43–49 CrossRef Medline
65. Hu M, Qian J, Pan S, et al. **Advancing medical imaging with language models: featuring a spotlight on ChatGPT.** *Phys Med Biol* 2024;69:10TR01 CrossRef
66. Mirzadeh I, Alizadeh K, Shahrokhi H, et al. **GSM-symbolic: understanding the limitations of mathematical reasoning in large language models.** *arXiv preprint arXiv:241005229* 2024
67. Bhayana R. **Chatbots and large language models in radiology: a practical primer for clinical and research applications.** *Radiology* 2024;310:e232756 CrossRef Medline
68. Bender EM, Gebru T, McMillan-Major A, et al. **On the dangers of stochastic parrots: Can language models be too big?** *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*; 2021:610–23
69. Obermeyer Z, Powers B, Vogeli C, et al. **Dissecting racial bias in an algorithm used to manage the health of populations.** *Science* 2019;366:447–53 CrossRef Medline
70. Ong JCL, Chang SY-H, William W, et al. **Medical ethics of large language models in medicine.** *NEJM AI* 2024;1:A1ra2400038 CrossRef
71. Li H, Moon JT, Purkayastha S, et al. **Ethics of large language models in medicine and medical research.** *Lancet Digit Health* 2023;5:e333–35 CrossRef Medline
72. Mannarswamy S, Chidambaram S. **Opening the NLP Blackbox—analysis and evaluation of NLP models: methods, challenges and opportunities.** *Proceedings of the 3rd ACM India Joint International Conference on Data Science & Management of Data (8th ACM IKDD CODS & 26th COMAD)*; 2021:447–448
73. Ajwani R, Javaji SR, Rudzicz F, et al. **LLM-generated black-box explanations can be adversarially helpful.** *arXiv preprint arXiv:240506800* 2024
74. Eaton P, Ince B, Iskova A. **The limitations of LLMs, or why are we doing RAG?** Accessed May 21, 2025. <https://www.enterprisedb.com/blog/limitations-llm-or-why-are-we-doing-rag>
75. Qiu S, Liu Q, Zhou S, et al. **Review of artificial intelligence adversarial attack and defense technologies.** *App Sci (Basel)* 2019;9:909 CrossRef
76. Zhang S, Metaxas D. **On the challenges and perspectives of foundation models for medical image analysis.** *Med Image Anal* 2024; 91:102996 CrossRef Medline
77. Chen L, Zaharia M, Zou J. **How is ChatGPT’s behavior changing over time?** *arXiv preprint arXiv:230709009* 2023
78. Delile J, Mukherjee S, Van Pamel A, et al. **Graph-based retriever captures the long tail of biomedical knowledge.** *arXiv preprint arXiv:240212352* 2024
79. Stokel-Walker C. **ChatGPT listed as author on research papers: many scientists disapprove.** *Nature* 2023;613:620–21 CrossRef Medline
80. Moy L. **Guidelines for use of large language models by authors, reviewers, and editors: considerations for imaging journals.** *Radiology* 2023;309:e239024 CrossRef Medline
81. Warraich HJ, Tazbaz T, Califf RM. **FDA perspective on the regulation of artificial intelligence in health care and biomedicine.** *J Am Med Assoc* 2024;333:241–47 CrossRef Medline
82. Wang G, Zhang S, Huang X, et al. **Editorial for special issue on explainable and generalizable deep learning methods for medical image computing.** *Med Image Anal* 2022;84:102727 CrossRef Medline
83. Kim W. **Seeing the unseen: advancing generative AI research in radiology.** *Radiology* 2024;311:e240935 CrossRef Medline
84. Marks M, Haupt CE. **AI chatbots, health privacy, and challenges to HIPAA compliance.** *J Am Med Assoc* 2023;330:309–10 CrossRef Medline
85. Akinci D’Antonoli T, Bluethgen C. **A new era of text mining in radiology with privacy-preserving LLMs.** *Radiol Artif Intell* 2024;6:e240261 CrossRef
86. Esmradi A, Yip DW, Chan CF. **A comprehensive survey of attack techniques, implementation, and mitigation strategies in large language models.** *International Conference on Ubiquitous Security*: Springer-Verlag; 2023:76–95
87. Surla A, Bodhankar A, Varshney T. **Generative AI: an easy introduction to multimodal retrieval-augmented generation.** March 20, 2024. Accessed February 28, 2025. <https://developer.nvidia.com/blog/an-easy-introduction-to-multimodal-retrieval-augmented-generation/>
88. Arasteh ST, Lotfinia M, Bressem K, et al. **RadioRAG: factual large language models for enhanced diagnostics in radiology using dynamic retrieval augmented generation.** *arXiv preprint arXiv:240715621* 2024
89. Open AI. **Introducing OpenAI o1-preview. A new series of reasoning models for solving hard problems.** September 12, 2024. Updated September 24, 2024. Accessed February 24, 2025. <https://openai.com/index/introducing-openai-o1-preview/>
90. Xexéo G, Braida F, Parreiras M, et al. **The economic implications of large language model selection on earnings and return on investment: a decision theoretic model.** *arXiv preprint arXiv:240517637* 2024
91. Jabal MS, Warman P, Zhang J, et al. **Language models and retrieval augmented generation for automated structured data extraction from diagnostic reports.** *arXiv preprint arXiv:240910576* 2024

# Neuroimaging Spectrum of Erdheim-Chester Disease: An Image-Based Review

Pranjal Rai, Hayden J. Swartz,  Neetu Soni,  John C. Benson,  Amit Agarwal,  Steven A. Messina,  Paul J. Farnsworth,  Carrie M. Carr, and  Girish Bathla



## ABSTRACT

**SUMMARY:** Erdheim-Chester disease (ECD) is a rare, multisystem histiocytic disorder characterized by its variable clinical presentations. CNS involvement is observed in approximately one-half of patients with ECD (up to 76% in some series) and often carries a poorer prognosis. While CNS involvement may remain asymptomatic, others may experience a range of neurologic symptoms, including cognitive decline, neuropsychiatric disturbances, motor deficits, cranial or peripheral neuropathies, and endocrine abnormalities. Neuroimaging findings in CNS-ECD are diverse, including neurodegeneration manifesting as cerebral or cerebellar volume loss; solitary or multifocal variably enhancing intraparenchymal lesions along the neuroaxis; meningeal infiltration; and involvement of the hypothalamo-pituitary axis, perivascular sheathing, or basal ganglia lesions. Other well-documented sites of involvement include the craniofacial region, orbits, and spine. Awareness of these findings is relevant, not only because of the nonspecific nature of these findings, but also because of the high proportion of CNS involvement in ECD and the higher mortality associated with CNS involvement. This review provides an in-depth overview of the various manifestations of CNS involvement in ECD and their imaging features, along with a brief overview of the differential considerations, which include other histiocytic and nonhistiocytic processes.

**ABBREVIATIONS:** CE = contrast-enhanced; CLIPPERS = chronic lymphocytic inflammation with pontine perivascular enhancement responsive to steroids; ECD = Erdheim-Chester disease; GWP = granulomatosis with polyangiitis; HPA = hypothalamo-pituitary axis; LCH = Langerhans cell histiocytosis; MAPK = mitogen-activated protein kinase; MEK = mitogen-activated protein kinase kinase; RDD = Rosai-Dorfman disease

Erdheim-Chester disease (ECD) is a rare non-Langerhans cell histiocytosis, first described in 1930 by 2 pathologists, Jacob Erdheim and William Chester.<sup>1</sup> While previously classified as an inflammatory disorder, it was reclassified as a hematopoietic neoplasm according to the 2016 World Health Organization classification. The revised 2016 classification of histiocytosis by the Histiocyte Society proposes 5 subtypes or categories of histiocytosis (L, C, R, M, and H) depending on characteristics that are clinical, radiologic, pathologic, phenotypic, genetic, and molecular. Out of these categories, both ECD and Langerhans cell histiocytosis (LCH) are included in the “L” (Langerhans) group, and Rosai-Dorfman disease (RDD) in the R group.<sup>2</sup> Even though CNS involvement in histiocytic disorders can have overlapping imaging manifestations, CNS involvement is

uncommon in RDD (<5%) and is predominantly confined to the hypothalamic-pituitary region (92.9%) in LCH.<sup>3-6</sup> The current work, therefore, is focused on imaging findings in ECD in the CNS, unless stated otherwise.

The exact prevalence of ECD remains undetermined. However, approximately 1500 cases have been documented in the English literature,<sup>7</sup> with patients typically presenting in the fifth decade of life.<sup>8</sup> The interval between symptom onset and diagnosis can range from several months to as long as 25 years, often attributed to misdiagnosis due to the absence of a specific clinical syndrome and the rarity of the disease itself.<sup>9</sup> Men are more frequently affected than women. The disorder can have a wide spectrum of clinical presentations, ranging from clinically indolent to life-threatening, potentially fatal disease. Patients with ECD have a median survival of 13.5 years and a 5-year survival rate of 82.7%.<sup>3</sup>

ECD is a multisystemic disorder and can involve virtually any organ system. The long bones of the upper and lower extremities are, however, most commonly affected, manifesting as symmetric medullary sclerosis.<sup>10</sup> Though CNS involvement is relatively common with systemic disease, isolated CNS disease is uncommon and primarily reported in case studies.<sup>3,11</sup> Notably, CNS involvement is an independent adverse prognostic factor and is associated with higher age at disease onset when compared with those without CNS involvement (median, 62 [range, 23–76] years versus 45 [range, 25–70] years;  $P = .03$ ).<sup>12</sup>

Received September 12, 2024; accepted after revision November 7.

From the Department of Radiology (P.R.), Tata Memorial Hospital, Maharashtra, India; Department of Radiology (H.J.S., J.C.B., S.A.M., P.J.F., C.M.C., G.B.), Division of Neuroradiology, Mayo Clinic, Rochester, Minnesota; and Department of Radiology (N.S., A.A.), Division of Neuroradiology, Mayo Clinic, Jacksonville, Florida.

Please address correspondence to Dr. Neetu Soni, Department of Radiology, Mayo Clinic, 4500 San Pablo Rd, Jacksonville, FL 55902; e-mail: Soni.neetu@mayo.edu; @pranj\_rad; @tmhradiology; @gbathlamd; @amitagarwalmd; @neetuso27437480; @mayoradiology

 Indicates open access to non-subscribers at [www.ajnr.org](http://www.ajnr.org)

 Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8599>

**Table 1: Comparison between the CNS manifestations of the common histiocytic disorders**

	Erdheim-Chester Disease	Langerhans Cell Histiocytosis	Rosai-Dorfman Disease
Demographics	M > F, sixth decade of life	M > F, first 2 decades of life	M > F, second to third decade of life
<i>BRAF</i> V600E mutation	Positive (more than LCD)	Positive	N/A
Histopathology	Foamy histiocytes, CD68+, CD1a-	Mononucleated dendritic cells, CD1a+, Birbeck granules	Multinucleated histiocytes, CD68+, S-100+, CD1a-; may be associated with immunoglobulin G4-related disease
Neurodegeneration and atrophy	+	++	+
Intraparenchymal lesions	Present (often multiple, more edema)	Present (often solitary, lesser edema)	Present (often multiple, more edema)
HPA involvement	+	+++	Rare
Extra-axial lesions	++	+	+
Craniofacial involvement	Sclerotic lesions (calvarial involvement is less common)	Lytic lesions (calvarial involvement is more common)	Rare
Orbital involvement	++	+	+
Vascular involvement	+	-	-

**Note:**—M indicates male; F, female; N/A, not applicable.

This review article explores the common and rare imaging manifestations of ECD within the CNS and discusses notable imaging differentials and cues that may help in accurate diagnosis. Given the diagnostic challenges and protean imaging phenotypes of CNS-ECD, a thorough understanding of imaging findings and imaging mimics may be helpful for timely and accurate diagnosis.

### **PATHOPHYSIOLOGY AND HISTOPATHOLOGIC FEATURES**

ECD is a neoplastic clonal myeloid form of histiocytosis, driven by mutations in the mitogen-activated protein kinase (MAPK) and phosphatidylinositol 3-kinase (PI3K)/protein kinase B (AKT) signaling pathways. Mutations affecting the MAPK signaling pathway are present in over 80% of patients with ECD.<sup>9</sup> Of these, *BRAF*V600E mutation is the most common (57%–70%), followed by *MAP2K1* mutations (20%).<sup>13</sup> The identification of *BRAF* and other MAPK pathway abnormalities, along with the observed co-occurrence of ECD and LCH in 15% of patients with histiocytosis, were important determinants of the 2016 reclassification of histiocytosis.<sup>2</sup>

Mutations in the MAPK pathway result in uninhibited proliferation of histiocytes, which are positive for CD68, CD163, Factor XIIIa, and Fascin, negative for CD1a and Langerin/CD207 (which are positively seen in LCH) with a variable expression of S-100.<sup>14</sup> Histologically, ECD is characterized by infiltrates of large, lipid-laden histiocytes (xanthoma cells) within a chronic inflammatory background of lymphocytes and plasma cells, with occasional Touton giant cells and fibrosis.<sup>15</sup> The identification of *BRAF* mutations offers the potential for targeted therapy with FDA-approved agents like Vemurafenib.<sup>16</sup> In patients with wild-type *BRAF*, treatment regimens incorporating mitogen-activated protein kinase kinase (MEK) inhibitors such as cobimetinib have demonstrated promising outcomes.<sup>17</sup>

### **CLINICAL FINDINGS**

ECD presents with a diverse range of clinical manifestations that vary in frequency but can impact both disease diagnosis and prognostication. In many instances, the disease has an indolent course and may be discovered incidentally during evaluation for another condition. The most common symptom is bone pain, as bone involvement occurs in approximately 95% of cases.<sup>18</sup> Bone

involvement is characterized by bilateral symmetric medullary sclerosis of the meta-diaphysis of long bones in the appendicular skeleton, with less frequent involvement of the skull and axial bones. This pattern of sclerosis contrasts with LCH, where lytic lesions are more common.<sup>19</sup> Table 1 summarizes the differences between CNS manifestations of common histiocytic disorders.

Patients with ECD exhibiting neurologic symptoms also often have concurrent bone lesions. Most patients with CNS-ECD are symptomatic, with asymptomatic involvement only seen in about 11%.<sup>3</sup> Neurologic manifestations vary, based on the site and severity of involvement, and can include cognitive impairment, neuropsychiatric disturbances, motor deficits, cerebellar dysfunction, cranial or peripheral neuropathies, seizures, oculomotor abnormalities, and myelopathy. Cognitive difficulties and neuropathies are the most commonly reported symptoms.<sup>20</sup> Additional symptoms indirectly associated with CNS involvement include headaches (secondary to meningeal involvement), diabetes insipidus (pituitary involvement), and exophthalmos (orbital involvement).<sup>21</sup>

A higher prevalence of the *BRAF*V600E mutation has been observed within the subset of patients with ECD with CNS involvement (60%–77%).<sup>3,4,22</sup> A systematic review by Cives et al<sup>23</sup> revealed that patients with CNS involvement were less likely to exhibit cardiac involvement, albeit nonsignificantly, and had a higher incidence of bone, skin, retroperitoneal, lung, aortic, and renal infiltration. Furthermore, over 50% of these patients experienced simultaneous involvement of at least 2 distinct anatomic sites within the CNS.<sup>23</sup> CNS involvement has also been independently identified as a predictor of both mortality and refractoriness to first-line therapies. Consequently, early detection of ECD in the CNS can potentially reduce the morbidity associated with ineffective immunotherapies and treatments.<sup>24</sup>

Additional systemic symptoms may also be seen, depending on the site and severity of involvement of the various organ systems. Myeloproliferative disorders and myelodysplastic syndromes are also seen in approximately 10% of ECD cases.<sup>25</sup>

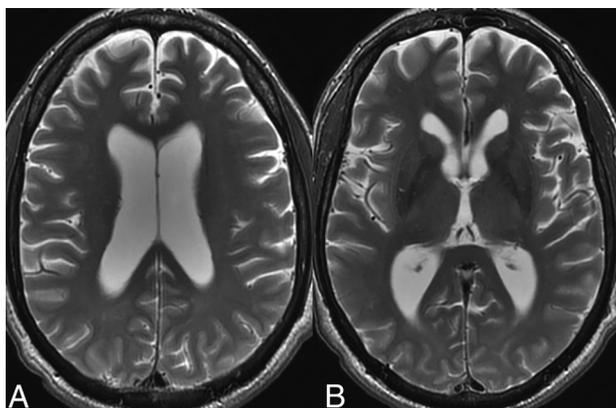
### **NEUROIMAGING FEATURES**

#### **Neurodegeneration**

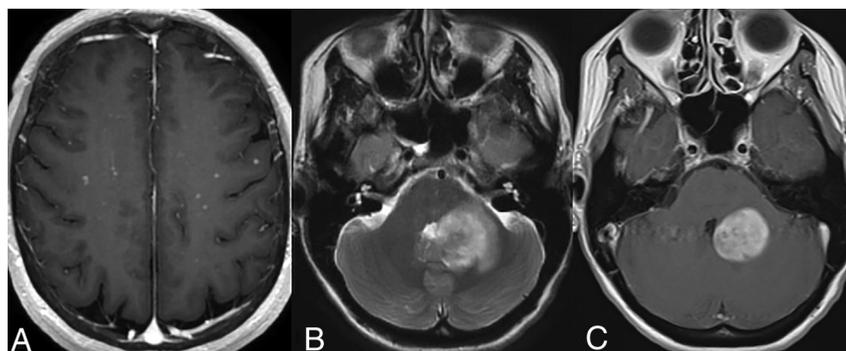
Prevalence of parenchymal volume loss (Fig 1) as a manifestation of underlying ECD has been reported in approximately 15% of

patients (over 60% in some studies).<sup>3,4,26</sup> Notably, most studies examining this relationship have relied on subjective or ordinal assessments rather than quantitative volumetric measurements. The underlying pathophysiology is hypothesized to be uncontrolled systemic inflammation resulting in elevated levels of proinflammatory cytokines, which can readily traverse the blood-brain barrier. The resulting inflammatory cascade and microglial activation results in neuronal loss and disruption of myelin integrity, ultimately leading to cortical and cerebellar atrophy. The term “pseudodegenerative” has been used by some authors because areas of axonal degeneration and myelin loss may not be accompanied by any histiocytic infiltration.<sup>3,27</sup>

Diamond et al<sup>28</sup> conducted a volumetric analysis in a small cohort ( $n=11$ ) of patients with ECD without CNS tumors or prior neurotoxic therapies, demonstrating diffuse reductions in cortical thickness and subcortical white matter. Their findings suggest that cognitive decline and behavioral changes in patients with ECD may, in part, be secondary to reduction in the brain volume. Even though there were no statistically significant differences in white matter or cerebellar volumes between patients with ECD and age-matched controls, the former showed a statistically significant reduction in cortical thickness and subcortical gray matter volume. Bhatia et al<sup>22</sup> observed disproportionate cerebellar atrophy compared with cerebral hemisphere volume loss



**FIG 1.** Parenchymal atrophy. Axial T2-weighted images in a 55-year-old patient with ECD demonstrate diffuse cortical atrophy with mild ex vacuo dilation of the lateral ventricles. There were no focal lesions.



**FIG 2.** Parenchymal lesions. T1-CE maximum intensity projection images in a patient (A) reveal scattered micronodular enhancing lesions bilaterally. Axial T2 WI (B) and T1-CE images (C) in a different patient demonstrate a masslike T2 hyperintense periventricular lesion with homogeneous enhancement and mild surrounding edema.

in 14% of their study population. Similar findings were observed in the study by Zahergivar et al<sup>4</sup> where infratentorial (cerebellar and brainstem) atrophy was observed more commonly than cerebral (20.7% versus 13.8%), indicating that in a subset of patients, there may be differential involvement of the infratentorial brain.<sup>4,22</sup> The latter also attempted to establish a correlation between cerebral atrophy and *BRAFV600E* mutation. While their findings narrowly missed statistical significance ( $P = .053$ ), the observed trend suggests the need for further research to investigate this potential association.<sup>4</sup>

Parenchymal volume loss may also be seen with other histiocytic disorders and may be proportionately more common with LCH. Fan et al,<sup>26</sup> for example, compared neuroimaging findings between LCH, ECD, and RDD and noted that both LCH and ECD exhibited degenerative patterns, with cortical atrophy being more prevalent in LCH.<sup>3,6</sup>

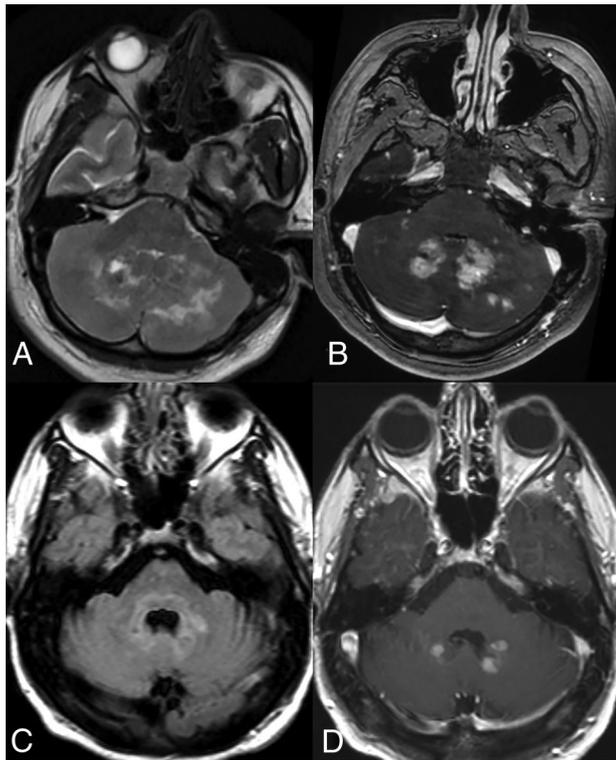
**Parenchymal Lesions.** These may involve the supra- or infratentorial brain. While initial studies found a predilection of these lesions for the posterior fossa,<sup>29,30</sup> Boyd et al,<sup>20</sup> in their prospective study, found a broad neurologic involvement throughout the craniospinal axis.

Lesions in the supratentorial compartment (seen in approximately 46% of cases) are encountered more frequently in the frontal lobe, followed by the parietal and occipital lobes (Fig 2). Intraparenchymal enhancing lesions can show considerable variation in size and were classified by Fan et al<sup>26</sup> into micronodular (<3 mm), nodular (3–10 mm), and masslike (>10 mm) subtypes. These lesions can be solitary to multiple, sometimes resembling metastatic disease.<sup>31</sup> On imaging, they typically exhibit T2 prolongation without any restricted diffusion, often have heterogeneous enhancement, and show minimal associated edema and mass effect.<sup>20</sup> On perfusion imaging, the intraparenchymal lesions demonstrate increased  $K^{trans}$  values, either with or without increased  $V_p$  values.<sup>22</sup>

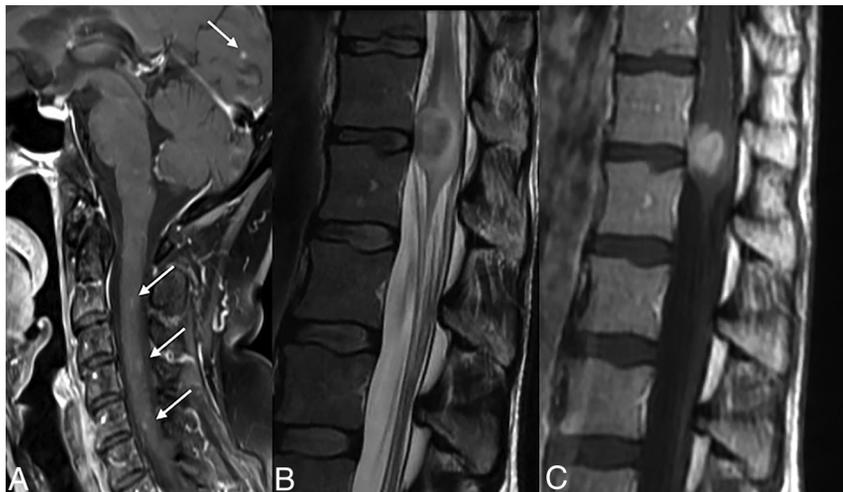
White matter involvement is variably described (60%–87%) and manifests as nondiffusion restricting supra- and infratentorial lesions with or without underlying enhancement. There is a lack of clarity about these lesions, with some authors considering them as a component of the neurodegenerative changes,<sup>3,26,32</sup> and others classifying them as parenchymal lesions attributable to ECD,<sup>33,34</sup> and the remainder not explicitly labeling these into either subtype.<sup>30</sup> Without a histopathologic diagnosis of these lesions, which is generally not performed in all cases, the exact etiology and definitive diagnosis remain uncertain and offer a further scope for research. Given that ECD is more commonly seen in patients in the fifth to seventh decade of life, it is conceivable that at least some of the non-enhancing lesions may overlap with underlying leukoaraiosis. Within the deep gray structures, basal ganglia involvement may occur in about 7% (most commonly putamen), followed by the amygdala and thalamus.<sup>4</sup> Less frequently described imaging findings

in the basal ganglia involvement include multiple punctate areas of signal loss on SWI in a pattern atypical of senile calcifications (Supplemental Data).<sup>35</sup> However, the precise implication of these findings remains unclear and warrants further investigation.

Infratentorial compartment involvement (Fig 3) is present in approximately one-half of all cases, affecting both the brainstem and cerebellum, which may show volume loss and enhancing or



**FIG 3.** Cerebellar involvement. Axial T2-WI (A) and T1-CE (B) images in a patient demonstrate ill-defined T2-isointense lesions in bilateral cerebellar hemispheres, most prominently along the dentate nuclei. Axial FLAIR (C) and T1-CE (D) images in a different patient reveal more focal involvement of the bilateral dentate nuclei.



**FIG 4.** Spinal involvement. Sagittal T1-CE image (A) reveals tiny discrete enhancing foci involving the cervical cord and the occipital lobe. Sagittal T2 WI (B) and T1-CE images (C) in a different patient demonstrate a masslike involvement of the conus.

nonenhancing lesions. The *BRAFV600E* mutation is frequently associated with cerebellar involvement, often characterized by the presence of bilateral lesions and frequent involvement of the dentate nucleus (15%).<sup>4,15,29</sup> Involvement of the bilateral superior or middle cerebellar peduncles is more common, with the inferior cerebellar peduncles less frequently affected. Within the brainstem, the pons is commonly involved, followed by the medulla and the midbrain.<sup>4</sup> The affected structures may also exhibit atrophy over time.<sup>34</sup>

ECD involving the spine can manifest in a uni- or multifocal pattern (Fig 4). Imaging findings may mimic demyelinating or inflammatory lesions.<sup>30,34,36</sup> Due to the rarity of spinal involvement, however, there is debate whether routine spine screening is required in the absence of symptoms.<sup>22</sup>

Uncommon CNS-ECD manifestations include curvilinear enhancement along the ependymal lining of the lateral ventricle with extension into the lentiform nucleus (Supplemental Data), cystic lesions with septal or ring enhancement, pineal or choroid plexus masses, and diffuse T2 hyperintensity of the pons with or without irregular transversely oriented enhancement.<sup>15</sup> Occasionally, punctate-enhancing lesions similar to chronic lymphocytic inflammation with pontine perivascular enhancement responsive to steroids (CLIPPERS) may also occur (Fig 5).<sup>26</sup>

Intraparenchymal lesions within any part of the neuraxis may exhibit prolonged gadolinium contrast retention, with some demonstrating a T1-shortening effect days after initial imaging.<sup>37,38</sup> The exact mechanism underlying this finding remains unclear, but it is hypothesized to be attributed to the abnormal retention of gadolinium by the histiocytes.<sup>39</sup>

**Hypothalamo-Pituitary Axis (HPA) Lesions.** HPA involvement is observed in approximately 17%–44%.<sup>3,4,22,30</sup> The most common imaging manifestation is the loss of posterior pituitary bright spot followed by pituitary stalk thickening.<sup>40</sup> Other findings include enhancing nodular or micronodular lesions involving the sella or infundibulum (Fig 6), deviation of the stalk, pituitary atrophy, or T2 hyperintensity within the hypothalamus.<sup>10,26,30,32</sup>

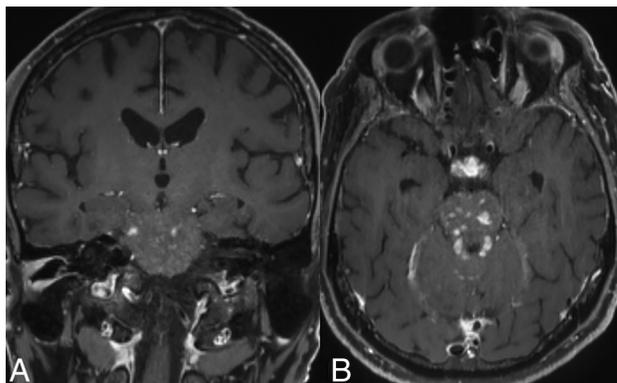
Diabetes insipidus is seen in nearly 50% of patients with CNS-ECD, secondary to HPA involvement.<sup>41</sup> Larger lesions may also present with anosmia or features of hypopituitarism. Endocrine disorders and HPA involvement are more common in LCH as compared with ECD.<sup>26</sup>

**Vascular Involvement.** Vascular involvement is seen in approximately 10%–17% of the cases and is unique to ECD compared with LCH and RDD.<sup>10,26</sup> When present, it is associated with a poor prognosis and higher mortality as the disease progression ultimately leads to fibrosis and persistent, progressive ischemic manifestations.<sup>26,32</sup>

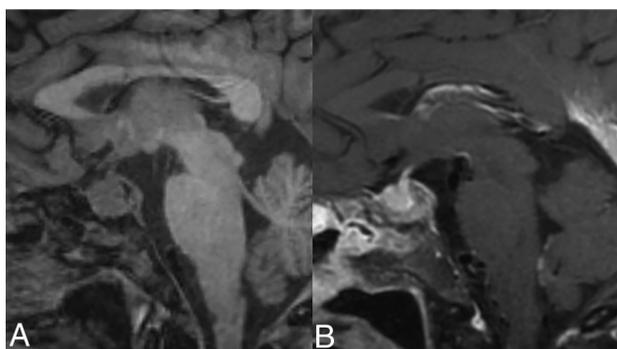
Direct vascular involvement in ECD occurs because of periadventitial vessel inflammation, in contrast to the transmural inflammation seen in other

vasculitides like Takayasu arteritis. This inflammation can lead to a secondary mass effect and compression of adjacent structures.

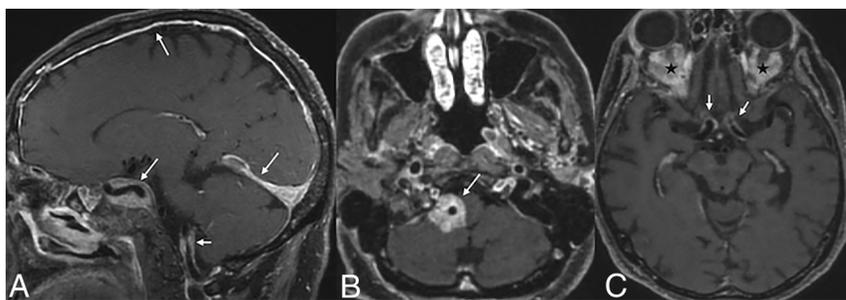
The vascular sheathing may exhibit a circumferential or non-circumferential segmental pattern of T2 isointense to hypointense, perivascular soft tissue with homogeneous enhancement



**FIG 5.** Coronal (A) and axial (B) TI-CE images in a patient with ECD demonstrate punctate areas of enhancement relatively confined to the pons, mimicking CLIPPERS. Note the thickening and enhancement within the infundibulum in (B) suggesting involvement of the HPA axis by the disease.



**FIG 6.** HPA axis involvement. Sagittal T1 noncontrast image (A) demonstrates an enlarged pituitary gland with loss of the posterior pituitary bright spot. Also, note abnormal marrow signal in clivus from bony involvement. Sagittal TI-CE image (B) demonstrates heterogeneous enhancement.



**FIG 7.** Vascular involvement. Sagittal (A) and axial (B and C) TI-CE in 3 different patients with ECD showing vascular involvement. There is vertebral artery involvement (*short arrow*, A) along with dural, tentorial, and cavernous sinus involvement (*long arrows*, A). In the second patient (B), there is vascular sheathing and enhancement around the right vertebral artery (*arrow*, B). The third image (C) shows similar involvement of the ICA vessels (*arrows*) along with orbital infiltrative lesions (*black stars*).

(Fig 7), predominantly involving the vessels of the posterior fossa such as the basilar trunk or the vertebral arteries.<sup>42</sup> Involvement of these or any of the vessels of the anterior circulation may also result in secondary stroke.<sup>43</sup> Vascular involvement in the neck, near the carotid bulb, may present as dysautonomia.<sup>20</sup> Vessel wall imaging can be helpful in indeterminate cases.

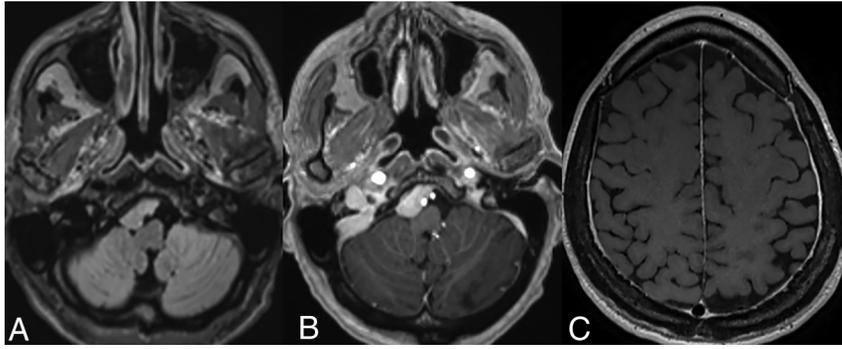
ECD presenting as dural masses may also secondarily involve the dural venous sinuses (Supplemental Data), with the sagittal, transverse, or straight sinuses being commonly affected, leading to thrombosis.<sup>10,44</sup> With pericarotid vascular sheathing, secondary extension into the cavernous sinuses (Supplemental Data) may be observed.<sup>10</sup> The presentation may be chronic, and the lesions tend to involve the venous sinuses focally without any intraluminal extension. Secondary venous infarcts, however, have not been reported in the literature.

**Extra-Axial Lesions.** Dura-based lesions, seen in approximately 30%–50% of cases,<sup>8</sup> generally involve the supratentorial compartment in the form of single or multiple dural-based masses. They are generally T2 hypointense, and invariably enhance on post-contrast sequences.<sup>22,40</sup> Similar to the parenchymal lesions, they frequently demonstrate increased  $K^{trans}$  values on perfusion imaging, either with or without increased  $V_p$  values.<sup>22</sup> While their imaging characteristics may resemble a meningioma,<sup>32</sup> concurrent parenchymal lesions or systemic disease favors the former, while adjacent bony hyperostosis and calcifications within the mass favor the latter. Another reported imaging finding is the presence of radiating T1- and T2-weighted hypointense spicules originating from the center of the mass, which do not exhibit enhancement.<sup>15</sup> Solitary fibrous tumors may also exhibit somewhat similar findings, characterized by a “yin-yang” appearance with varying degrees of low and high T2 signal intensity within different areas of the lesion. This can make it challenging to differentiate from extra-axial lesions associated with ECD, especially in the absence of intraparenchymal lesions or systemic disease.<sup>45</sup>

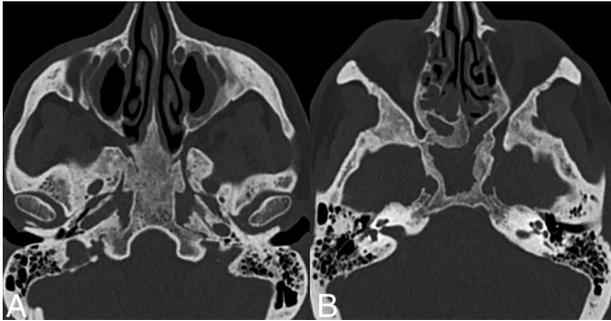
The presence of dural masses makes patients with CNS-ECD susceptible to spontaneous atraumatic subdural hematomas.<sup>46</sup> Extra-axial involvement can also manifest as diffuse pachymeningeal thickening (Fig 8) and enhancement over cerebral or cerebellar convexities.<sup>47</sup>

Spinal extradural involvement with ECD may present with compressive myelopathy (Supplemental Data). Lesions may also mimic schwannomas when involving the exiting nerve roots.<sup>48</sup> Patients may occasionally have concurrent paraspinal muscle involvement, presenting as enhancing, infiltrative masses. However, the lack of any pathognomonic imaging findings in ECD makes it challenging to accurately diagnose the condition, especially without a comprehensive clinical context.

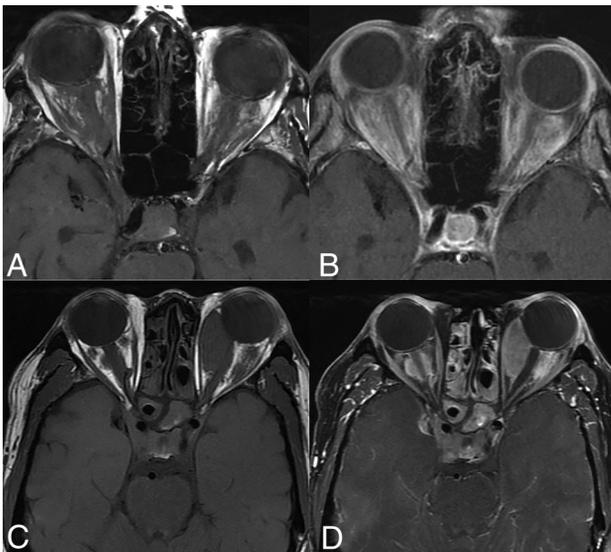
Leptomeningeal disease in ECD is uncommon (6%–7%) and may manifest as smooth or nodular leptomeningeal



**FIG 8.** Axial FLAIR (A) and T1-CE (B) images demonstrate an enhancing dural-based lesion along the clivus. A T1-CE (C) image of another patient reveals diffuse pachymeningeal enhancement.



**FIG 9.** Sinonasal ECD. Axial CT images (bone kernel) in a patient with ECD demonstrate skull base and sinus involvement characterized by thickening and osteosclerosis of the basisphenoid and paranasal sinuses.



**FIG 10.** Orbital involvement in ECD. Axial T1 WI precontrast (A and C) and T1-CE (B and D) images of 2 different patients showing infiltrative (A and B) and masslike (C and D) orbital involvement. There is involvement of the bilateral cavernous sinuses in both cases along with paranasal sinus involvement in the bottom row (C and D).

enhancement (Supplemental Data).<sup>8,49</sup> The presence of meningeal disease may not necessarily be accompanied by signs of clinical meningitis. CSF analysis generally demonstrates elevation

of protein, decrease in glucose levels, and rarely, the presence of foamy histiocytes.<sup>49,50</sup>

**Craniofacial Involvement.** Craniofacial involvement in ECD is seen in 40%–50% of cases and manifests as discrete or confluent osteosclerotic lesions with variable bony thickening (Fig 9). The most frequently affected bone is the calvaria, followed by the skull base and paranasal sinuses. On MRI, they typically appear hypointense on both T1- and T2-weighted images.<sup>3,22,40</sup>

Calvarial involvement in ECD is less frequent than in other histiocytosis (such as LCH). It is frequently accompanied by concurrent involvement of other skeletal sites, such as the ribs and long bones of the appendicular skeleton.<sup>30,51</sup> For patients with potentially referable neurologic symptomatology, a targeted CT scan of the sinuses and/or skull base may help find causative osseous lesions.<sup>22</sup>

**Orbital Involvement.** Orbital involvement in ECD is typically bilateral (18%–30%), with the intraconal compartment most commonly involved (Fig 10).<sup>3,18,40</sup> Both infiltrative and masslike involvement may occur and present with T2WI hypointensity and enhancement. The resulting mass effect and compression of the optic nerve can lead to proptosis, visual deficits, retro-orbital pain, and oculomotor nerve palsy.<sup>20,21</sup> In some cases, the lesions may extend into the extraconal compartment to secondarily involve the lacrimal gland.<sup>52</sup> Involvement of the optic chiasm, however, is uncommon.<sup>20</sup> It is also notable that orbital involvement is higher than in LCH and RDD, which may help differentiate between these entities.<sup>26</sup>

Primary optic nerve involvement, characterized by optic nerve sheath enlargement or signal abnormalities, is observed in approximately 47% of cases.<sup>53</sup> Choroidal involvement, though rare in ECD, may present as a masslike intraorbital lesion.<sup>54</sup> Table 2 summarizes the common neuroimaging manifestations of ECD.

### POSTTREATMENT IMAGING FINDINGS IN ECD

A contrast-enhanced (CE) MRI of the brain with gadolinium is recommended for all patients with ECD at the time of diagnosis.<sup>21</sup> Regular follow-up imaging is crucial for patients with CNS involvement undergoing treatment, preferably more closely during the early stages.<sup>34</sup> Treatment options include surgical resection, radiation therapy, high-dose corticosteroids, targeted agents, and other systemic therapies.<sup>22</sup>

There is a limited body of literature on the posttreatment imaging features of CNS-ECD. Many patients with CNS involvement undergo debulking surgery for extra-axial masses or receive *BRAF*/*MEK* inhibitors due to their rapid response.<sup>47</sup> Post treatment, lesions often demonstrate a reduction in size and enhancement (Supplemental Data), though residual deficits may persist in patients with nonenhancing lesions. Surrounding edema may also show interval improvement.<sup>15,55</sup> In some cases, near-complete remission of imaging findings

**Table 2: Summary of the CNS imaging findings in Erdheim-Chester disease (reported prevalence in parentheses)**

Compartment	Imaging Finding
CNS involvement (25%–76%)	Parenchymal volume loss (cortical and cerebellar atrophy), increased radial diffusivity of water molecules
Parenchyma (supratentorial [approx. 46%]; posterior fossa [20%–46%]; white matter [60%–87%]; basal ganglia [7%])	Scattered lesions (micronodular, nodular, masslike), T2 prolongation, heterogeneous enhancement, minimal edema, increased $K^{trans}$ values T2/FLAIR hyperintense lesions of indeterminate significance May have signal drop-out on SWI in basal ganglia Prolonged gadolinium contrast retention
HPA (17%–44%)	Thickening of pituitary stalk, hypothalamic T2 hyperintensity Loss of posterior pituitary bright spot, nodular or micronodular lesions, pituitary atrophy
Vascular (10%–17%)	Periadventitial vessel inflammation, vascular sheathing Dural venous sinus thrombosis or infiltration
Extra-axial (dural [30%–50%], leptomeningeal [6%–7%])	Dura-based solitary or multiple lesions (enhancing, isointense), pachymeningeal thickening, spinal extradural involvement Leptomeningeal disease
Craniofacial (40%–50%)	Osteosclerotic lesions (calvaria, skull base, paranasal sinuses), bone thickening
Orbital (18%–30%)	Intraconal or extraconal masses Optic nerve sheath enlargement Choroidal involvement (intraorbital mass)

may be seen.<sup>47</sup> Cohen Aubart et al<sup>3</sup> noted that MRI findings are generally concordant with disease activity, and it is rare to have worsening of imaging findings in patients who are clinically improving or stable. In general, however, clinical status is a more reliable indicator of treatment effectiveness.<sup>22</sup>

### DIFFERENTIAL DIAGNOSIS

Given the multiple and often nonspecific imaging manifestations, CNS-ECD can have a broad differential diagnosis. The closest differential considerations are other histiocytosis subtypes (eg, LCH and RDD), and histopathologic examination may be the only definitive means of differentiation. In their comparative study, Fan et al<sup>26</sup> noted that CNS-LCH lesions are often solitary and invariably involved the hypothalamic-pituitary region. RDD, on the other hand, may have overlapping manifestations with CNS-ECD but is considerably rarer and involves a younger age group. Vascular involvement was exclusively seen with CNS-ECD. Similarly, bony lesions are rare in RDD, lytic in LCH, and invariably sclerotic in ECD.

Brain metastases may present as enhancing parenchymal lesions but are considerably more common, often cortical in location, and may have associated hemorrhage or cystic change. Vascular sheathing, CLIPPER like pattern, T2 shortening, and concurrent bilateral craniofacial involvement would be atypical. Similarly, hypothalamic-pituitary involvement may occur but is generally uncommon. A whole-body PET-CT scan can aid in identifying the primary tumor site.<sup>56</sup> In equivocal cases, a stereotactic biopsy may be warranted. Leptomeningeal carcinomatosis may also mimic CNS-ECD but can be differentiated based on CSF cytology and biochemical findings.<sup>57</sup>

Patients with primary CNS lymphoma are generally older (unless immunocompromised) and more likely to show diffusion restriction in lesions. Involvement of the HPA axis and dura is less common, as is parenchymal atrophy in untreated patients.<sup>58</sup> Neurosarcoidosis, on the other hand, can have several overlapping imaging manifestations in terms of parenchymal, dural, and meningeal involvement. However, the bony craniofacial and orbital involvement is less common and more focal when present.

Dilated medullary veins, as noted in a subset of patients with neurosarcoidosis, are not present in CNS-ECD.<sup>59,60</sup> Spinal involvement in sarcoidosis generally involves  $\geq 3$  vertebral segments.<sup>61</sup> Patchy brainstem involvement can also occur in the context of CLIPPERS or intravascular primary CNS lymphomas, but these conditions generally show a dramatic response to steroids, unlike ECD.<sup>62</sup>

CNS-ECD and granulomatosis with polyangiitis (GWP, previously Wegner granulomatosis) both may show meningeal, sinus, and orbital involvement. However, GWP typically shows osteolytic lesions with nasal septal erosion and frequently shows extraconal and unilateral orbital involvement, while long-bone involvement is rare.<sup>63</sup>

CNS IgG4-related disease predominantly affects the pituitary gland and meninges rather than the parenchyma and cranial nerves. Hypothalamic-pituitary axis involvement can manifest as thickening or infiltration of the infundibulum or enlargement of the pituitary gland. The most common site of orbital involvement is the lacrimal gland followed by the extraocular muscles or intra- or extraconal orbital fat.<sup>64,65</sup> Orbital lesions may resemble pseudotumors on imaging, with definitive diagnosis often requiring biopsy for confirmation.

### CONCLUSIONS

While the imaging features of CNS-ECD may overlap with other diseases, a high index of clinical suspicion and a pattern-recognition approach can aid in diagnosis. Given that CNS involvement is noted in most cases, it is essential that neuroradiologists are familiar with the diverse imaging manifestations of this condition. The association of CNS-ECD with a poorer prognosis and a higher incidence of *BRAFV600E* mutations further underscores the importance of accurate diagnosis and targeted therapy (with *BRAF* or *MEK* inhibitors). Awareness of the neuroimaging spectrum of ECD can play a crucial role in early detection and guiding treatment decisions.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

- Matzumura M, Arias-Stella J, Novak JE. **Erdheim-Chester disease: a rare presentation of a rare disease.** *J Investig Med High Impact Case Rep* 2016;4:2324709616663233 CrossRef Medline
- Emile JF, Abia O, Fraitag S; Histiocyte Society, et al. **Revised classification of histiocytoses and neoplasms of the macrophage-dendritic cell lineages.** *Blood* 2016;127:2672–81 CrossRef Medline
- Cohen Aubart F, Idbaih A, Galanaud D, et al. **Central nervous system involvement in Erdheim-Chester disease: an observational cohort study.** *Neurology* 2020;95:e2746–54 CrossRef Medline
- Zahergivar A, Firouzabadi FD, Homayounieh F, et al. **Central nervous system involvement in Erdheim-Chester disease: a magnetic resonance imaging study.** *Clin Imaging* 2024;115:110281 CrossRef Medline
- Vaidya T, Mahajan A, Rane S. **Multimodality imaging manifestations of Rosai-Dorfman disease.** *Acta Radiology Open* 2020;9:2058460120946719 CrossRef Medline
- Grois N, Fahrner B, Arceci RJ; Histiocyte Society CNS LCH Study Group, et al. **Central nervous system disease in Langerhans cell histiocytosis.** *J Pediatr* 2010;156:873–81 CrossRef Medline
- Histiocytosis Association. **Erdheim-Chester disease.** November 12, 2020. Accessed August 25, 2024. <https://histio.org/histiocytic-disorders/erdheim-chester-disease/>
- Haque A, Pérez CA, Reddy TA, et al. **Erdheim-Chester disease with isolated CNS involvement: a systematic review of the literature.** *Neurol Int* 2022;14:716–26 CrossRef Medline
- Haroche J, Cohen-Aubart F, Amoura Z. **Erdheim-Chester disease.** *Blood* 2020;135:1311–18 CrossRef Medline
- Drier A, Haroche J, Savatovsky J, et al. **Cerebral, facial, and orbital involvement in Erdheim-Chester disease: CT and MR imaging findings.** *Radiology* 2010;255:586–94 CrossRef Medline
- Garg N, Lavi ES. **Clinical and neuroimaging manifestations of Erdheim-Chester Disease: a review.** *J Neuroimaging* 2021;31:35–44 CrossRef Medline
- Toya T, Ogura M, Toyama K, et al. **Prognostic factors of Erdheim-Chester disease: a nationwide survey in Japan.** *Haematologica* 2018;103:1815–24 CrossRef Medline
- Diamond EL, Durham BH, Haroche J, et al. **Diverse and targetable kinase alterations drive histiocytic neoplasms.** *Cancer Discov* 2016;6:154–65 CrossRef Medline
- Fasulo S, Alkomos MF, Pjetergjoka R, et al. **Erdheim-Chester disease presenting at the central nervous system.** *Autops Case Rep* 2021;11:e2021321 CrossRef Medline
- Sedrak P, Ketonen L, Hou P, et al. **Erdheim-Chester disease of the central nervous system: new manifestations of a rare disease.** *AJNR Am J Neuroradiol* 2011;32:2126–31 CrossRef Medline
- Fernández-Eulate G, Muñoz-Lopetegui A, Ruiz I, et al. **Vemurafenib as first-line therapy in BRAF-V600E-mutant Erdheim-Chester disease with CNS involvement.** *BMJ Case Rep* 2019;12:e228280 CrossRef
- Diamond EL, Durham BH, Ulaner GA, et al. **Efficacy of MEK inhibition in patients with histiocytic neoplasms.** *Nature* 2019;567:521–2 CrossRef Medline
- Estrada-Veras JI, O'Brien KJ, Boyd LC, et al. **The clinical spectrum of Erdheim-Chester disease: an observational cohort study.** *Blood Adv* 2017;1:357–66 CrossRef Medline
- Ambrosini V, Savelli F, Merli E, et al. **F-18 FDG PET/CT detects muscle involvement in Erdheim-Chester disease.** *Clin Nucl Med* 2012;37:196–97 CrossRef Medline
- Boyd LC, O'Brien KJ, Ozkaya N, et al. **Neurological manifestations of Erdheim-Chester disease.** *Ann Clin Transl Neurol* 2020;7:497–506 CrossRef Medline
- Goyal G, Heaney ML, Collin M, et al. **Erdheim-Chester disease: consensus recommendations for evaluation, diagnosis, and treatment in the molecular era.** *Blood* 2020;135:1929–45 CrossRef Medline
- Bhatia A, Hatzoglou V, Ulaner G, et al. **Neurologic and oncologic features of Erdheim-Chester disease: a 30-patient series.** *Neuro Oncol* 2020;22:979–92 CrossRef Medline
- Cives M, Simone V, Rizzo FM, et al. **Erdheim-Chester disease: a systematic review.** *Crit Rev Oncol Hematol* 2015;95:1–11 CrossRef Medline
- Arnaud L, Hervier B, Néel A, et al. **CNS involvement and treatment with interferon- $\alpha$  are independent prognostic factors in Erdheim-Chester disease: a multicenter survival analysis of 53 patients.** *Blood* 2011;117:2778–82 CrossRef Medline
- Papo M, Diamond EL, Cohen-Aubart F, et al. **High prevalence of myeloid neoplasms in adults with non-Langerhans cell histiocytosis.** *Blood* 2017;130:1007–13 CrossRef Medline
- Fan X, Liu T, Zhang Z, et al. **Comparison of neuroimaging features of histiocytic neoplasms with central nervous system involvement: a retrospective study of 121 adult patients.** *Eur Radiology* 2023;33:8031–42 CrossRef Medline
- Perry VH. **Contribution of systemic inflammation to chronic neurodegeneration.** *Acta Neuropathol* 2010;120:277–86 CrossRef Medline
- Diamond EL, Hatzoglou V, Patel S, et al. **Diffuse reduction of cerebral grey matter volumes in Erdheim-Chester disease.** *Orphanet J Rare Dis* 2016;11:109 CrossRef Medline
- Kumar P, Singh A, Gamanagatti S, et al. **Imaging findings in Erdheim-Chester disease: what every radiologist needs to know.** *Pol J Radiology* 2018;83:e54–62 CrossRef Medline
- Parks NE, Goyal G, Go RS, et al. **Neuroradiologic manifestations of Erdheim-Chester disease.** *Neurol Clin Pract* 2018;8:15–20 CrossRef Medline
- Wang F, Cao X, Niu N, et al. **Multisystemic imaging findings in Chinese patients with Erdheim-Chester disease.** *AJR Am J Roentgenol* 2019;213:1179–86 CrossRef Medline
- Rangankar V, Ajmera P, Agarwal N, et al. **Erdheim-Chester disease: with neurological manifestation and multisystem involvement: case report and radiological review.** *Egypt J Radiol Nucl Med* 2021;52:261 CrossRef
- Wang Y, Camelo-Piragua S, Abdullah A, et al. **Neuroimaging features of CNS histiocytosis syndromes.** *Clin Imaging* 2020;60:131–40 CrossRef Medline
- Benson JC, Vaubel R, Ebne BA, et al. **Erdheim-Chester disease.** *AJNR Am J Neuroradiol* 2023;44:505–10 CrossRef Medline
- Mazor RD, Manevich-Mazor M, Kesler A, et al. **Clinical considerations and key issues in the management of patients with Erdheim-Chester disease: a seven case series.** *BMC Med* 2014;12:221 CrossRef Medline
- Jeon I, Choi JH. **Isolated thoracic intramedullary Erdheim-Chester disease presenting with paraplegia: a case report and literature review.** *BMC Musculoskelet Disord* 2021;22:270 CrossRef Medline
- Tien RD, Brasch RC, Jackson DE, et al. **Cerebral Erdheim-Chester disease: persistent enhancement with Gd-DTPA on MR images.** *Radiology* 1989;172:791–92 CrossRef Medline
- Babu RP, Lansan TA, Chadburn A, et al. **Erdheim-Chester disease of the central nervous system: report of two cases.** *J Neurosurg* 1997;86:888–92 CrossRef Medline
- Adem C, Hélie O, Lévêque C, et al. **Case 78: Erdheim-Chester disease with central nervous system involvement.** *Radiology* 2005;234:111–15 CrossRef Medline
- Aswani Y, Patel A, Zhan X, et al. **Imaging in Erdheim-Chester disease.** *RadioGraphics* 2024;44:e240011 CrossRef Medline
- Conley A, Manjila S, Guan H, et al. **Non-Langerhans cell histiocytosis with isolated CNS involvement: an unusual variant of Erdheim-Chester disease.** *Neuropathology* 2010;30:634–47 CrossRef Medline
- Suzuki H, Wanibuchi M, Komatsu K, et al. **Erdheim-Chester disease involving the central nervous system with the unique appearance of a coated vertebral artery.** *NMC Case Rep J* 2016;3:125–28 CrossRef Medline
- Takeuchi T, Sato M, Sonomura T, et al. **Erdheim-Chester disease associated with intramedullary spinal cord lesion.** *Br J Radiology* 2012;85:e62–64 CrossRef Medline
- Viswanathan S, Kadir NA, Lip AC, et al. **Central nervous system Erdheim-Chester disease presenting with raised intracranial pressure and cerebellar signs mimicking neurosarcoidosis with secondary**

- cerebral venous thrombosis. *Neurol India* 2014;62:446–48 CrossRef Medline
45. Weon YC, Kim EY, Kim HJ, et al. **Intracranial solitary fibrous tumors: imaging findings in 6 consecutive patients.** *AJNR Am J Neuroradiol* 2007;28:1466–69 CrossRef Medline
  46. Noh SM, Kang HG. **Embolic infarction with subdural hemorrhage in Erdheim-Chester disease.** *J Clin Neurol* 2020;16:349–51 CrossRef Medline
  47. Pan Z, Kleinschmidt-DeMasters BK. **CNS Erdheim-Chester disease: a challenge to diagnose.** *J Neuropathol Exp Neurol* 2017;76:986–96 CrossRef Medline
  48. Huang Z, Li S, Hong J, et al. **Erdheim-Chester disease mimicking lumbar nerve schwannoma: case report and literature review.** *Spinal Cord Ser Cases* 2019;5:90 CrossRef Medline
  49. Polk C, Weida C, Patel N, et al. **Erdheim-Chester disease presenting as meningitis with hypoglycorrachia: A case report.** *Medicine (Baltimore)* 2022;101:e30585 CrossRef Medline
  50. Globerman H, Burstein S, Girardina PJ, et al. **A xanthogranulomatous histiocytosis in a child presenting with short stature.** *Am J Pediatr Hematol Oncol* 1991;13:42–46 CrossRef Medline
  51. Caparros-Lefebvre D, Pruvo JP, Rémy M, et al. **Neuroradiologic aspects of Chester-Erdheim disease.** *AJNR Am J Neuroradiol* 1995;16:735–40 Medline
  52. De Abreu MR, Chung CB, Biswal S, et al. **Erdheim-Chester disease: MR imaging, anatomic, and histopathologic correlation of orbital involvement.** *AJNR Am J Neuroradiol* 2004;25:627–30 Medline
  53. Haroche J, Gueniche Y, Galanaud D, et al. **Erdheim-Chester disease: look it in the eye. An orbital magnetic resonance imaging study.** *Haematologica* 2022;107:2667–74 CrossRef Medline
  54. Abdellatif A, Mason CM, Ytterberg SR, et al. **Choroidal involvement in Erdheim-Chester disease.** *Ophthalmic Surg Lasers Imaging Retina* 2015;46:674–76 CrossRef Medline
  55. Wang JN, Qiu Y, Niu N, et al. **Successful treatment of central nervous system involved Erdheim-Chester disease by intermediate-dose cytarabine as first-line therapy.** *Acta Oncol* 2020;59:302–05 CrossRef Medline
  56. Grazzini I, Venezia D, Del Roscio D, et al. **Morphological and functional neuroradiology of brain metastases.** *Semin Ultrasound CT MR* 2023;44:170–93 CrossRef Medline
  57. Bönig L, Möhn N, Ahlbrecht J, et al. **Leptomeningeal metastasis: the role of cerebrospinal fluid diagnostics.** *Front Neurol* 2019;10:839 CrossRef Medline
  58. Bathla G, Hegde A. **Lymphomatous involvement of the central nervous system.** *Clin Radiology* 2016;71:602–09 CrossRef Medline
  59. Bathla G, Abdel-Wahed L, Agarwal A, et al. **Vascular involvement in neurosarcoidosis: early experiences from intracranial vessel wall imaging.** *Neurol Neuroimmunol Neuroinflammation* 2021;8:e1063 CrossRef
  60. Bathla G, Singh AK, Policeni B, et al. **Imaging of neurosarcoidosis: common, uncommon, and rare.** *Clin Radiol* 2016;71:96–106 CrossRef Medline
  61. Soni N, Bathla G, Pillenahalli Maheshwarappa R. **Imaging findings in spinal sarcoidosis: a report of 18 cases and review of the current literature.** *Neuroradiol J* 2019;32:17–28 CrossRef Medline
  62. Berkman J, Ford C, Johnson E, et al. **Misdiagnosis: CNS Erdheim-Chester disease mimicking CLIPPERS.** *Neuroradiol J* 2018;31:399–402 CrossRef Medline
  63. Guzman-Soto MI, Kimura Y, Romero-Sanchez G, et al. **From head to toe: granulomatosis with polyangiitis.** *Radiographics* 2021;41:1973–91 CrossRef Medline
  64. Huynh KN, Kong MJ, Nguyen BD. **Anatomic and functional imaging of immunoglobulin G4-related disease and its mimics.** *RadioGraphics* 2023;43:e220097 CrossRef Medline
  65. Baptista B, Casian A, Gunawardena H, et al. **Neurological manifestations of IgG4-related disease.** *Curr Treat Options Neurol* 2017;19:14 CrossRef Medline

# Sodium MRI in Pediatric Brain Tumors

 Aashim Bhatia, Cassie Kline, Peter J. Madsen, Michael J. Fisher, Fernando E. Boada, and  Timothy P.L. Roberts



## ABSTRACT

**SUMMARY:** Direct sodium MRI ( $^{23}\text{Na}$ -MRI) derives its signal from spin-manipulation of the  $^{23}\text{Na}$  nucleus itself and not the more conventional and familiar  $^1\text{H}$ -MRI. Although present at much lower concentrations in the human body than the  $^1\text{H}$  nuclei in the water molecule  $\text{H}_2\text{O}$ , advances in coil design and pulse sequence development have enabled the feasibility of human in vivo  $^{23}\text{Na}$ -MRI. Additionally,  $^{23}\text{Na}$ -MRI has the potential to offer nuanced physiologic insights not available to conventional MRI; this feature forms the basis of interest in its development and optimism for its novel clinical utility.  $^{23}\text{Na}$ -MRI has the potential to be a useful noninvasive imaging technique to assess biochemical and physiologic cellular changes in tissues, eg, cell integrity and tissue viability. Pathologically, the concentration of total sodium is elevated in tumors relative to normal counterparts due to increased intracellular sodium and/or an increased proportion of extracellular space (reflecting changes in cell morphology and anomalies of homeostasis). Here we review the technological advancements with improved pulse sequences and reconstruction methods that counter the inherent challenges of measuring sodium concentrations in the pediatric brain (in particular, its short-tissue T2 value) and present detailed imaging approaches to quantifying sodium concentrations in the pediatric brain that can be assessed in various CNS pathologies, with the focus on pediatric brain tumors.

**ABBREVIATIONS:** BSC = bound sodium concentration; IBR = iterative Bowsher reconstruction; IR = inversion recovery; ISC = intracellular sodium concentration;  $^{23}\text{Na}$ -MRI = sodium MRI; RF = radiofrequency; TPI = twisted projection imaging; TSC = total sodium concentration

**D**irect sodium MRI ( $^{23}\text{Na}$ -MRI) exploits similar atomic nuclear spin mechanisms that allow conventional proton ( $^1\text{H}$ ) MRI but derives its signal from manipulation of the  $^{23}\text{Na}$  nucleus itself.<sup>1,2</sup> Although present at much lower concentrations in the human body than the  $^1\text{H}$  nuclei in the water molecule  $\text{H}_2\text{O}$ , advances in coil design and pulse sequence development have enabled the feasibility of human in vivo  $^{23}\text{Na}$ -MRI. This uses the same MR scanner as conventional MRI (though it must be equipped with “heteronuclear” capability, primarily a broadband radiofrequency [RF] system, since the Larmor frequency of the  $^{23}\text{Na}$  spins differs from that of  $^1\text{H}$  protons). For the same reason, a dedicated transmit/receive RF coil is required; however, these are now commercially available for use at both 3T and 7T. While

conventional pulse sequences (eg, gradient-recalled echo) can often be manipulated to function at the resonance frequency of sodium, 33.8 MHz at 3T, instead of the conventional 127.1 MHz for  $^1\text{H}$  (based on the gyromagnetic ratios of sodium and hydrogen), dedicated pulse sequences are often preferred and will be discussed below. Given the challenges with sensitivity, coil hardware, and pulse sequence requirements, the adoption of  $^{23}\text{Na}$ -MRI might present some initial challenges. However, it has the potential to offer nuanced physiologic insights not available to conventional MRI, which form the basis of interest in its development and optimism for its novel clinical utility.  $^{23}\text{Na}$ -MRI has the potential to be a useful noninvasive imaging technique to assess biochemical and physiologic cellular changes in tissues, eg, cell integrity and tissue viability.<sup>3,4</sup>

$^{23}\text{Na}$ -MRI has demonstrated such potential in the pathophysiologic evaluation of acute ischemic stroke, multiple sclerosis, amyotrophic lateral sclerosis, migraines, and multiple tumor types, including tumors both within and outside of the CNS.<sup>3,5-11</sup> There is no radiation exposure with MRI (including  $^{23}\text{Na}$ -MRI), and  $^{23}\text{Na}$ -MRI does not use gadolinium-based contrast agents, eliminating the potential long-term risks of these agents commonly used in  $^1\text{H}$ -MRI; nor does it, in fact, rely on any other exogenous tracer (in contrast to PET, for example). The sensitivity of direct sodium imaging in detecting disease states stems from the tightly controlled  $\text{Na}^+$  ion homeostasis in healthy tissues, which

Received October 1, 2024; accepted after revision December 6.

From the Department of Radiology (A.B., T.P.L.R.), Division of Oncology (C.K., M.J.F.), and Division of Neurosurgery (P.J.M.), Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; and Radiological Sciences Laboratory (F.E.B.), School of Medicine, Stanford, California.

This work was supported by an American Brain Tumor Association Discovery Grant, Society of Pediatric Radiology Pilot Grant, Department of Defense Exploration Grant.

Please address correspondence to Aashim Bhatia, MD, 3500 Civic Center Blvd, 2nd Floor, Office 2592 Philadelphia, PA 19104; e-mail: aashimbhatia@gmail.com; @draash

 Indicates open access to non-subscribers at [www.ajnr.org](http://www.ajnr.org)

 Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8642>

maintains a large concentration gradient between intracellular sodium concentration (ISC) at 10–15 mmol/L and extracellular sodium at ~145 mmol/L. The concentration of total sodium is elevated in tumor cells relative to normal tissues due to increased intracellular sodium concentration and/or an increased proportion of extracellular space (reflecting changes in cell morphology).<sup>8</sup>

In neoplastic tissue, sustained depolarization of the cell membrane with altered Na<sup>+</sup>/H<sup>+</sup> pump transport kinetics precedes the high rate of mitotic activity that characterizes abnormal tumor growth, leading to a concomitant increase in the ISC that has been demonstrated in a number of human neoplasms (Supplemental Data).<sup>12–14</sup> Further characterization of this rise in intracellular sodium concentration in several types of human carcinomas and glial cell lines has established a positive correlation between proliferative activity and an increased intracellular Na<sup>+</sup>:K<sup>+</sup> ratio (mostly due to an increase in ISC). An increase in ISC contributes to an increase in total sodium concentration (TSC). Research suggests that elevated abnormal changes in voltage-gated sodium channels (Na<sub>v</sub>1.3, Na<sub>v</sub>1.6) in low-grade astrocytomas and increased expression of the Na, K, -ATPase β-subunit isoforms in peritumoral astrocytes may contribute to the increases in both ISC and TSC.<sup>15</sup> In addition, pediatric brain tumor tissues consist of up to 30% non-neoplastic cells, including glioma-associated microglia, monocyte-derived macrophages, myeloid-derived suppressor cells, and tumor-infiltrating CD4<sup>+</sup> and CD8<sup>+</sup> T-cells.<sup>16</sup> The tumor microenvironment is similarly characterized by low glucose, low oxygen, and high acidity, resulting from aerobic glycolysis (Warburg effect) in highly proliferative cancer cells to meet their high nutrient demand, despite the presence of oxygen and fully functioning mitochondria.<sup>17,18</sup> Activation of other Na<sup>+</sup> ion transport proteins (such as H<sup>+</sup> extrusion protein Na<sup>+</sup>/H<sup>+</sup> pump and cell volume regulatory protein Na-K-Cl pumps) could also contribute to increased ISC and TSC in tumor tissues, which have been detected in glioma cells as well as in glioma-associated microglia and peritumoral reactive astrocytes.<sup>19</sup> Changes of Na<sup>+</sup> transport protein expression and function in tumor cells, tumor-associated glial cells, and tumor infiltrating immune cells play a role in the elevated sodium concentrations in pediatric brain tumors.<sup>20</sup> Thus, direct sodium imaging with <sup>23</sup>Na-MRI offers attractive sensitivity to the processes of the cell membrane.

There have been promising results suggesting that sodium levels can predict CNS tumor types and even predict progression-free survival.<sup>11,21–23</sup> Na-MRI has been shown to be a marker of tumor proliferation in animal glioma models and, most important, has shown the ability to distinguish between posttreatment necrosis and treatment responses.<sup>22</sup> In adult (human) brain tumors, <sup>23</sup>Na-MRI has demonstrated the ability to provide an additional marker for differentiating high-grade and low-grade

tumors, treatment response, and even determining a molecular mutation, with direct measurement of sodium concentrations complementary to <sup>1</sup>H-MRI.<sup>11,13,21</sup> In a study of 20 adult patients with malignant brain tumors, the mean sodium concentration (in millimoles per kilogram wet weight) was 61 (SD, 8) for gray matter, 69 (SD, 10) for white matter, 135 (SD, 10) for CSF, 113 (SD, 14) for vitreous humor, and, most important, 103 (SD, 36) for brain tumor, which is significantly elevated compared with reference brain tissues.<sup>12</sup> Recently, we published the first report of application of <sup>23</sup>Na-MRI to pediatric brain tumor imaging.<sup>13</sup>

Technical challenges in <sup>23</sup>Na-MRI in the pediatric brain necessitate an optimized MRI sequence, for which there is currently no standard approach or consensus. This article will summarize the multiple technical factors and hardware involved in estimating sodium concentrations using <sup>23</sup>Na-MRI in the pediatric brain, considering pulse sequences, RF coils, and reconstruction methods. In addition, incompletely appreciated physiology associated with sodium changes in the brain during myelination or even in neonates with relatively more incomplete myelination compared with older children present areas of interest in the research field as well as essential questions that need to be answered as a part of clinical implementation.

### Equipment

Details of the <sup>23</sup>Na-MRI technique will be dependent on the MRI system and hardware available. A brief overview of hardware requirements will be necessary before discussing the <sup>23</sup>Na-MRI technique. Sodium concentrations are 1000s of fold less than hydrogen protons in the human body, particularly in the human brain, where there is an average of 35–45 mmol/L of sodium ions compared with 80M proton concentration in the hydrogen nuclei of water molecules (Table).<sup>23</sup> Not only is the sodium ion concentration so much lower (compromising sensitivity), but the fundamental sensitivity of the nuclear magnetic resonance (NMR) experiment is intrinsically lower for the <sup>23</sup>Na nucleus compared with <sup>1</sup>H, being linearly dependent on the gyromagnetic ratio (which differs among nuclei and is approximately 25% for sodium compared with hydrogen). This finding compromises sensitivity (and, ultimately, SNR) further. Worse still, the T2-decay constants of sodium nuclei in vivo are very short (especially for the intracellular component—as low as ~3–5 ms at 3T), presenting a further challenge to signal detection, requiring very short TEs. These factors have generally led to a practical minimum field strength requirement of 3T to obtain adequate sodium signal. Additionally, because of the different Larmor frequency of sodium (approximately 33.8 MHz at 3T), the scanner will need to be equipped with a broadband RF system (sometimes called heteronuclear, or X-nucleus, package).

**Sensitivity of alternative nuclei in brain<sup>a</sup>**

Nucleus	Gyromagnetic Ratio MHz/T	Scanner Frequency	Tissue Concentration	Relative NMR Sensitivity	Relative Biologic Sensitivity
<sup>1</sup> H	42.58	127.1	88 mol/L	1	1
<sup>23</sup> Na	11.26	33.8	35–45 mmol/L	0.092	~0.0001
<sup>31</sup> P	17.24	51.7	1–10 mmol/L	0.0663	~0.00001
<sup>2</sup> D	6.54	19.5	0	0.0000096	0
<sup>19</sup> F	40.08	120.24	0	0.83	0

**Note:**—NMR indicates nuclear magnetic resonance.

<sup>a</sup> <sup>2</sup>D and <sup>19</sup>F are tracer techniques, as endogenous tissue concentrations are zero. Clinically, only <sup>23</sup>Na (and imaginably <sup>31</sup>P) provide sufficient in vivo sensitivity.

Since the US Food and Drug Administration (FDA) first approved the 3T system in 1999, research in  $^{23}\text{Na}$ -MRI has benefited from improved gradients and electronics to improving the clinical feasibility through shorter minimum TE values. Lower spatial resolution ( $\sim 3$  mm isotropic) is often an accepted compromise along with relatively long acquisition times ( $\sim 10$  minutes) due to the low SNR. SNR increases approximately linearly with increased field strengths (though T2 values may shorten further), and there are, thus, advantages of imaging at higher field strengths, namely 7T, which will be discussed below.

### Imaging Technique

TSC within the brain is a signal from both the extracellular and intracellular spaces. The extracellular space consists primarily of the interstitial space, which contains extracellular matrix and cells. Vascular (3% of volume) and perivascular spaces are also included in the extracellular space but occupy a much smaller percentage. The rapid  $^{23}\text{Na}$ -MRI signal loss, which was initially believed to be from primarily short-T2 intracellular compartments, is now considered reflective of both interstitial and intracellular spaces. Much interest is focused on resolving the intracellular and short-T2 component of the sodium signal, sometimes called the bound sodium concentration (BSC).

### Acquisition Approaches in $^{23}\text{Na}$ -MRI

Given intracellular sodium T2 values as short as  $\sim 3$ –5 ms compared with relatively longer (though still short) T2 values in the extracellular space ( $\sim 25$ –30 ms), conventional imaging sequences with relatively longer TEs will be preferentially sensitive to the extracellular and, even free-fluid, sodium signal only. Consequently, images acquired with conventional TEs  $> \sim 5$  ms tend to have the imaging characteristics of CSF maps, with relatively weak (if detectable) contribution from brain parenchyma.

Thus, ultra-short TEs ( $\sim 0.5$ –5 ms) are required to obtain signal from the rapidly decaying signal from sodium ions in the body (especially for sodium ions in the intracellular environment). Short-TE scanning can be achieved using some gradient-echo sequences; ultra-short TE scanning tends to use some variant of spiral or radial scanning (with signal sampling starting at or near the center of  $k$ -space) to maximize signal sensitivity.<sup>24-26</sup>

Even typically-used radial-scanning readout results in SNR loss and inefficient sampling of  $k$ -space. Optimized approaches to sampling  $k$ -space in  $^{23}\text{Na}$ -MRI have been developed, such as twisted projection imaging (TPI), which improves the SNR.<sup>27</sup> Ultra-short TE sequences acquire spin density-weighted  $^{23}\text{Na}$ -MRI (note these are  $^{23}\text{Na}$  spin density, and not “proton density”), which allows directly estimating sodium concentrations in tissues.

There is clinical interest in sodium concentrations that are reflective of the intracellular space, which can be more representative of pathologic diseases in children, such as in neoplasms, demyelinating disease (multiple sclerosis), and stroke.

Several strategies have been explored to suppress the relatively hyperintense  $^{23}\text{Na}$ -MRI signal from CSF (also, sometimes, necrosis) to emphasize relatively lower signal contributions from brain parenchyma. Inversion recovery (IR) (based on the difference in sodium T1 values) is an approach to attempt to suppress the sodium signal from the CSF; however, IR has not gained tremendous interest due

to the low SNR and limitations with specific absorption rate (SAR) due to  $180^\circ$  pulses.<sup>28</sup>

So-called dual echo sodium imaging has been performed in adults and children, in which 2 images are acquired (potentially in the same pulse sequence) at 2 different (but both short) TEs. The signal from these images is modeled, pixelwise, for a pool comprising 2 components. A simple linear equation then allows the separation of these component pools without incurring the SNR losses associated with the IR method, provided some mathematical assumptions are made. This approach allows a much greater SNR and lower SAR compared with IR, while still achieving  $^{23}\text{Na}$  fluid suppression. Since the dual-echo approach is essentially a subtraction technique, it is also vulnerable to differential sensitivity to magnetic susceptibility artifacts, leading to erroneous elevated signal on subtraction in regions of bone-tissue or air-tissue interface, which may masquerade as fluid-suppressed tissue sodium signal.<sup>27</sup>

To illustrate the dual-echo approach with corresponding mathematics, we describe a simplified model formulation (after Qian et al, 2015).<sup>27</sup>

Consider a simple 2-compartment model, consisting of 1 fraction  $\alpha_F$  in a “free” environment, characterized by the T2-decay constant  $T_{2F}$  and a second (“tissue”) fraction ( $= 1 - \alpha_F$ ) characterized by a different T2-decay constant,  $T_{2T}$ .

$$\text{FREE} : \alpha_F, T_{2F} \quad \text{TISSUE} (1 - \alpha_F), T_{2T}.$$

The observed sodium signal  $S$ , for a pulse sequence with TE, will be an average over the 2 pools (weighted by compartment volume fraction and T2 values):

$$S = S_0 \cdot \{ \alpha_F \cdot \exp(-TE/T_{2F}) + (1 - \alpha_F) \cdot \exp(-TE/T_{2T}) \},$$

noting that  $T_{2F}$  is relatively longer and  $T_{2T}$  is shorter. So, for 2 imaging pulse sequences with TEs 0.5 and 5 ms respectively:

$$\begin{aligned} \text{Short TE} : S(0.5) &= S_0 \cdot \{ \alpha_F \cdot \exp(-0.5/T_{2F}) \\ &+ (1 - \alpha_F) \cdot \exp(-0.5/T_{2T}) \}, \end{aligned}$$

$$\begin{aligned} \text{Long TE} : S(5.0) &= S_0 \cdot \{ \alpha_F \cdot \exp(-5/T_{2F}) \\ &+ (1 - \alpha_F) \cdot \exp(-5/T_{2T}) \}. \end{aligned}$$

Simple Subtraction (since the free component is relatively bright on both)

$$S(0.5) - S(5.0) \sim S_0 \cdot \{ (1 - \alpha_F) \cdot [\exp(-0.5/T_{2T}) - \exp(-5/T_{2T})] \}$$

$$\sim S_0 \cdot \{ (1 - \alpha_F) \cdot [\exp(-0.5/T_{2T}) - 0] \} \text{ if } T_{2T} < 5 \text{ ms}$$

$$\sim S_0 (1 - \alpha_F), \text{ if } T_{2T} > 0.5 \text{ ms,}$$

suggesting the so-called “short T2 image,” formed by simple subtraction, yields a signal proportional to the tissue fraction of sodium  $\alpha_T = (1 - \alpha_F)$ .

In practice, this simple subtraction fails to completely suppress longer T2 species, since the assumption that

$$\{\alpha_F \cdot \exp(-0.5/T_{2F})\} - \{\alpha_F \cdot \exp(-5/T_{2F})\} \sim 0,$$

is not met exactly for realistic values of  $T_{2F}$ . Consequently, we propose a weighted subtraction better suited for nulling of long  $T_{2F}$  species (eg, the fluid hyperintensities that obscure the conspicuity of pathology in the in vivo brain). Consider a weighting factor,  $\beta$ , (where  $\beta$  is approximately 15%):

Weighted Subtraction:

$$\begin{aligned} \text{"S}(0.5) - (1 + \beta)\text{S}(5.0)" \sim \text{So. } \{ & (1 - \alpha_F) \cdot [\exp(-0.5/T_{2T}) \\ & - (1 + \beta)\exp(-5/T_{2T})]\}. \end{aligned}$$

In this case, the free compartment has much reduced residual signal since the assumption,

$$\{\alpha_F \cdot \exp(-0.5/T_{2F})\} - (1 + \beta) \{\alpha_F \cdot \exp(-5/T_{2F})\} \sim 0,$$

does indeed hold; thus, the weighted subtraction image yields signal proportional to the tissue sodium fraction alone. The factor,  $\beta$ , needed to allow complete subtraction of the free pool signal depending on the T2 value of that fraction ( $T_{2F}$ ) and on the 2 chosen TEs. It can be determined empirically (during postprocessing) or estimated on the basis of the literature  $T_{2F}$  values. Thus, the weighted subtraction can be considered to reflect the tissue sodium fraction (weighted only by  $T_{2T}$  values). Illustrations of this methodology throughout this article will use a  $\beta = 15\%$  as an exemplar. Note, as with all subtraction techniques, but especially with those with relatively longer acquisition times (approximately 10 minutes), misregistration of the 2 image sets can lead to subtraction artifacts. A pulse sequence with interleaved TEs could be created or reregistration could be facilitated with the use of sodium chloride solution fiducials analogous to the commonly used vitamin E capsules for navigation sequences, to combat this outcome.

In reality, even the cellular (non-“free”) pool may itself be multicompartmental. This possibility leads to either a higher-order multicompartmental approach or at least a hierarchical multistep, 2-compartment model, likely necessitating further image acquisitions (with different TE choices). Such multicompartment models have been used to attempt to separate intracellular sodium from extracellular sodium concentrations. Combining fluid-suppressed IR and spin density-weighted imaging has also been used to estimate the pseudo-intracellular sodium concentration and extracellular volume fraction. The reliability of such a 3-compartment model has been challenged due to disease-related fluctuations in the parameters whose values are fixed by assumption: fixed relaxation times, fixed intracellular sodium concentration, and fixed volume fractions. Others have suggested a fourth compartment, myelin water, which contains equal ratios of intracellular and extracellular sodium.<sup>29</sup> The further expansion of differentiating the sodium signal into 3 or even possibly 4 compartments may, thus, be needed to improve the accuracy of the representation of the sodium physicochemical microenvironment. Nonetheless, the 2-compartment approach offers promise for ready clinical adoption.

At this point, the alternative strategies for sodium quantification based on triple quantum filtering may indeed offer more precise specificity but at the price of both SNR as well as analysis complexity and are considered beyond the scope of current clinical implementation.

### Localized T2\* Mapping Schemes

T2\* measured at multiple TEs has the potential to measure sodium concentrations in different molecular environments. Localized T2\* mapping can take a long time, but if it is combined with optimized multiecho readout schemes (3D MERINA), the acquisition times can be decreased.<sup>30</sup> Furthermore, local T2\* mapping may also provide a practical basis for improving the local image blurring associated with T2\*-dependent decay of signal along the spiral (or other)  $k$ -space trajectory.

Other factors in the correction of relaxation effects include TEs desired, which cannot be obtained due to hardware limitations, switching times between transmit and receive channels of the RF coil, or SAR limitations restricting the minimum pulse duration (because of its high peak  $B_1$  amplitude and, thus, RF power), and in certain tissues with very short transverse relaxation times, this leads to signal loss from T2\* relaxation. In the clinical setting, short TRs that do not allow complete longitudinal relaxation (<100 ms) are sometimes used. Such short TRs can lead to biases in the estimation of the tissue sodium concentration. If the T1 and T2\* relaxation times are known or can be reasonably estimated, relaxation effects can be corrected by calculating the relaxation-correction factors for mono and biexponential relaxation.

### Advanced Imaging Reconstruction

There are multiple approaches to image reconstructions to decrease the rather long acquisition times of  $^{23}\text{Na}$ -MRIs and/or improve the relatively low resolution. Radial and spiral acquisitions commonly used in  $^{23}\text{Na}$ -MRI can take advantage of compressed sensing for iterative reconstruction. With the increasing implementation of deep learning-based reconstruction algorithms in  $^1\text{H}$ -MRI, analogous optimization of  $^{23}\text{Na}$ -MRI is to be expected. However, the greater the degree of “actual” undersampling, the greater is the possibility of inaccurate estimation of the true sodium concentrations, which has been shown in prior studies.<sup>31</sup>

$^1\text{H}$ -MRI provides high-resolution, low-noise anatomic detail that can be combined with  $^{23}\text{Na}$ -MRI. Beyond simple overlay (as in eg, fMRI), this result might best be approached with a constrained reconstruction approach. This method considers the possibility of losing important sodium data, and only shared data between  $^1\text{H}$ - and  $^{23}\text{Na}$ -MRI are interpreted. This reconstruction approach allows improved resolution and low-noise sodium images with greater contrast between gray matter and white matter.<sup>32</sup> One emerging technique is the iterative Bowsher reconstruction (IBR) framework, which essentially deconvolves spatial T2-blurring using a T2 estimate from 2 TEs and anatomic priors from high-resolution T1-weighted or FLAIR images (Fig 1).<sup>32</sup> This method is widely used in sharpening PET images and has been adapted to  $^{23}\text{Na}$ -MRI. The proton MRI is coregistered with the  $^{23}\text{Na}$  MR images to confirm the location of brain tumors.

**Clinical Implementation Vignette.** We illustrate the above concepts with a specific and detailed example of a patient with a diffuse midline glioma of the pons. In the sodium MR imaging acquisition,  $^{23}\text{Na}$ -MR images were acquired on a 3T scanner (Magnetom Prisma; Siemens), equipped with heteronuclear capabilities; the RF coil for  $^{23}\text{Na}$  is a dual-tuned ( $^1\text{H}$ - $^{23}\text{Na}$ ) head volume coil (Rapid MRI). A custom-developed pulse sequence, TPI, is used to acquire  $^{23}\text{Na}$ -MRI data for TSC and BSC imaging using a dual-TE technique (~10 minutes per sequence, total of ~20 minutes to acquire both TEs), analogous to our prior work: field of view =

220 mm, matrix size =  $64 \times 64 \times 64$ , voxel size = 3.44 mm (isotropic),  $\text{TE}_1/\text{TE}_2 = 0.5/5$  ms,  $\text{TR} = 100$  ms, averages = 4, and Time to Acquisition (TA) = 10 minutes, 38 seconds for each TE.

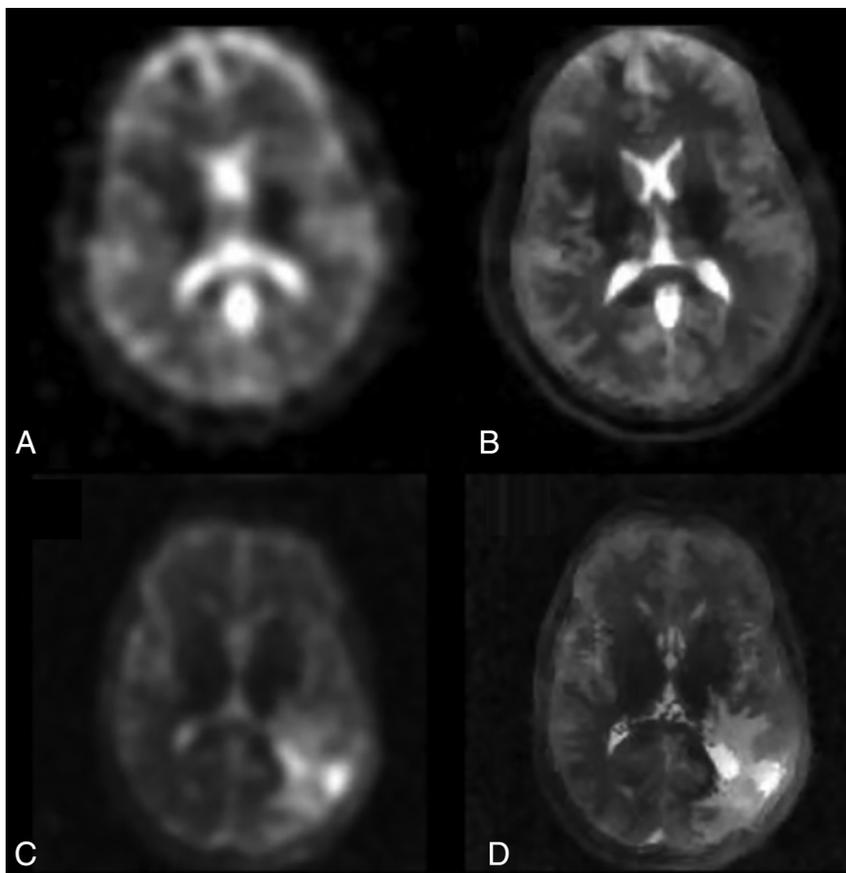
**TPI and Dual-Echo-Based Reconstruction.** The dual-echo TPI approach to measure BSC allows both an optimized ultra-short TE (as low as 0.3 ms) acquisition as well as a deblurred image reconstruction based on  $T_2^*$  fitting, rendering higher-quality reconstructed images (with reduced  $T_2^*$ -blurring). The ultra-short TE is achieved via asymmetric sampling of projection imaging with radiations from the center of the  $k$ -space along the polar and azimuthal angles. The TE is only limited by the RF excitation pulse and the associated rephasing gradient in case of slice selection. The dual-TE  $^{23}\text{Na}$ -MRI allows suppression of the elevated “free” sodium signal within both CSF and necrotic foci, resulting in improved distinction of non-neoplastic from neoplastic tissue. In the examples below (Figs 2–3), we use the 115%-weighted subtraction described above to optimize free fluid sodium suppression.

The TE is only limited by the RF excitation pulse and the associated rephasing gradient in case of slice selection. The dual-TE  $^{23}\text{Na}$ -MRI allows suppression of the elevated “free” sodium signal within both CSF and necrotic foci, resulting in improved distinction of non-neoplastic from neoplastic tissue. In the examples below (Figs 2–3), we use the 115%-weighted subtraction described above to optimize free fluid sodium suppression.

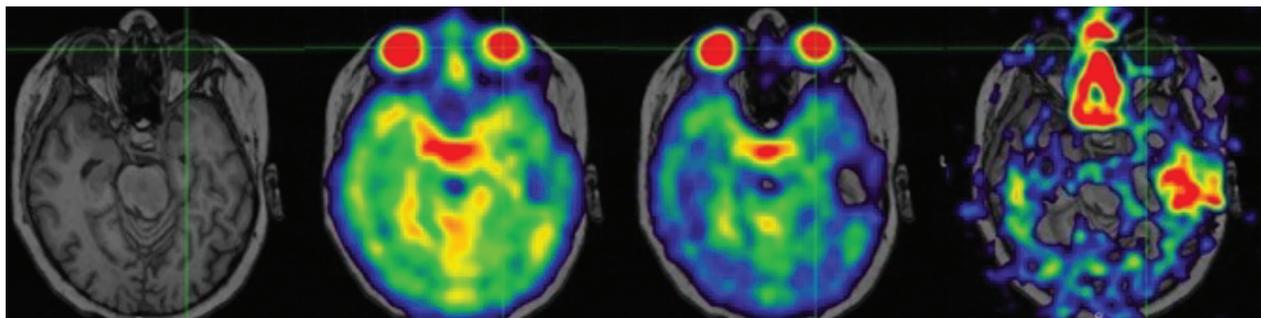
**$^{23}\text{Na}$  Image Reconstruction: Postprocessing and Quantitative Sodium MR Imaging**

We have developed a postprocessing method that uses the intensity of  $\text{TE}_1$  to linearly calibrate using the CSF region ( $\text{TSC} = 145$  mmol/L) and noise-only background ( $\text{TSC} = 0$  mmol/L). Sodium images are registered to anatomic T1-weighted, T2-weighted, or FLAIR proton images using 6 *df* rigid body transformation in medical imaging processing, analysis, and visualization (MIPAV; (<https://mipav.cit.nih.gov/>) software (Fig 4).

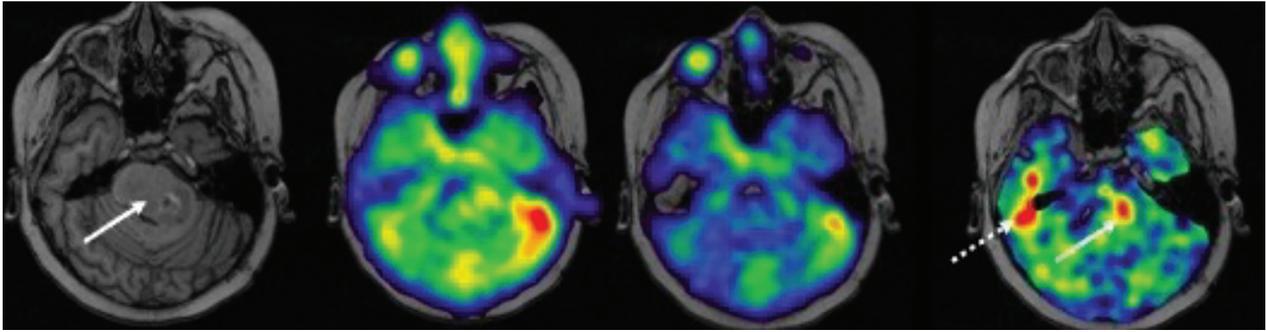
Imaging reconstruction methods based on similar low-resolution PET and fMRI have been applied to  $^{23}\text{Na}$ -MRI with promising results. Incorporating



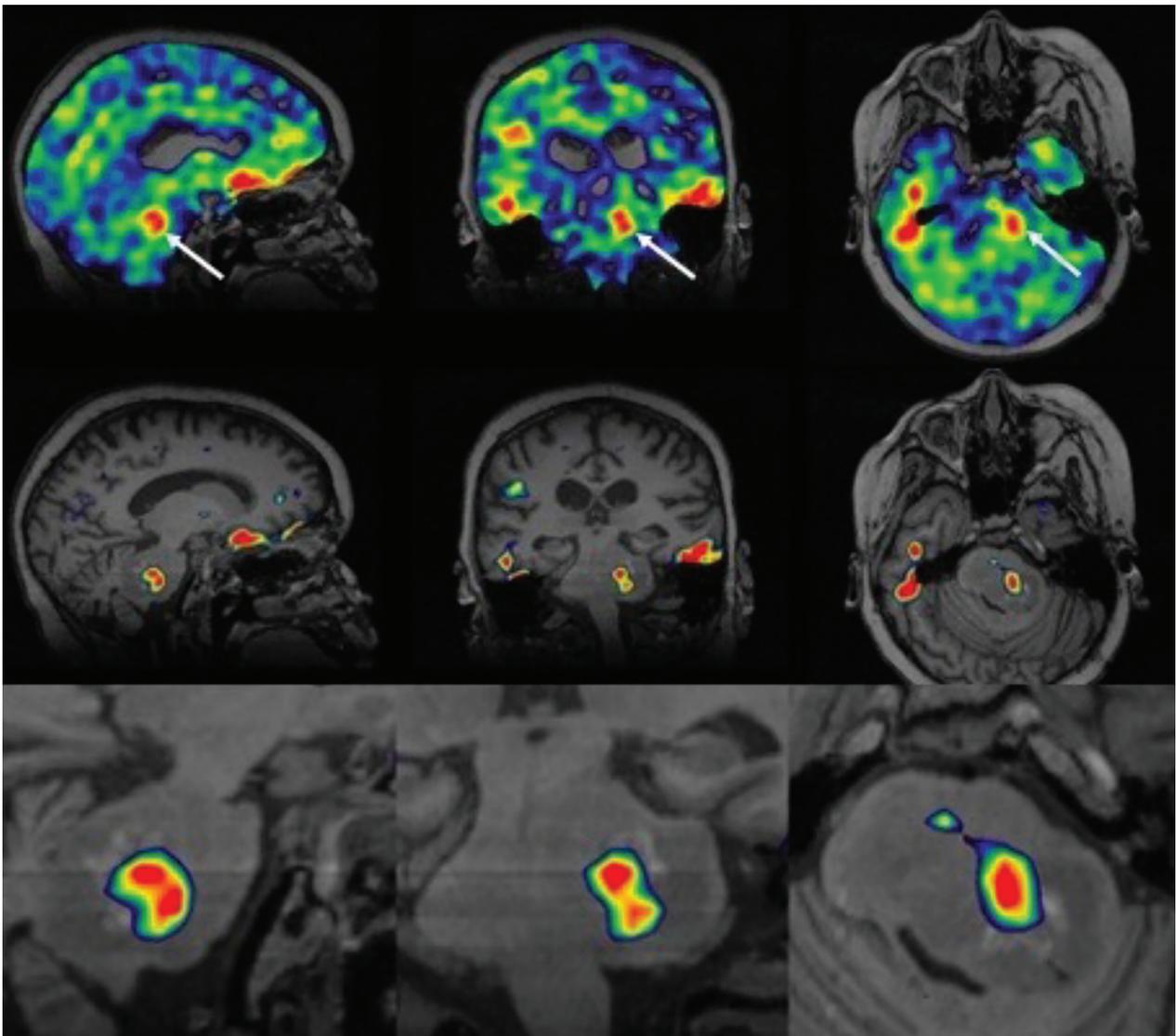
**FIG 1.** Deblurring  $^{23}\text{Na}$ -MRI with iterative Bowsher reconstruction. A, Native  $^{23}\text{Na}$ -MRI of a healthy volunteer (A) can be significantly sharpened to B using 2-point  $T_2$  estimation and anatomic priors. Similarly, C and D demonstrate the benefit of IBR deblurring in a patient with glioma, allowing clearer intralesional resolution.



**FIG 2.** Dual-TE  $^{23}\text{Na}$ -MRI in a pediatric patient with suppression of the elevated “free” sodium signal within vitreous fluid of the globe (cross-hair). Note  $T_2$ -based signal loss from  $\text{TE} = 0.5$ –5 ms. Also note, susceptibility artifacts in the left temporal lobe and nasal cavity, becoming especially pronounced on subtraction. (Same patient in Figs 3–6).



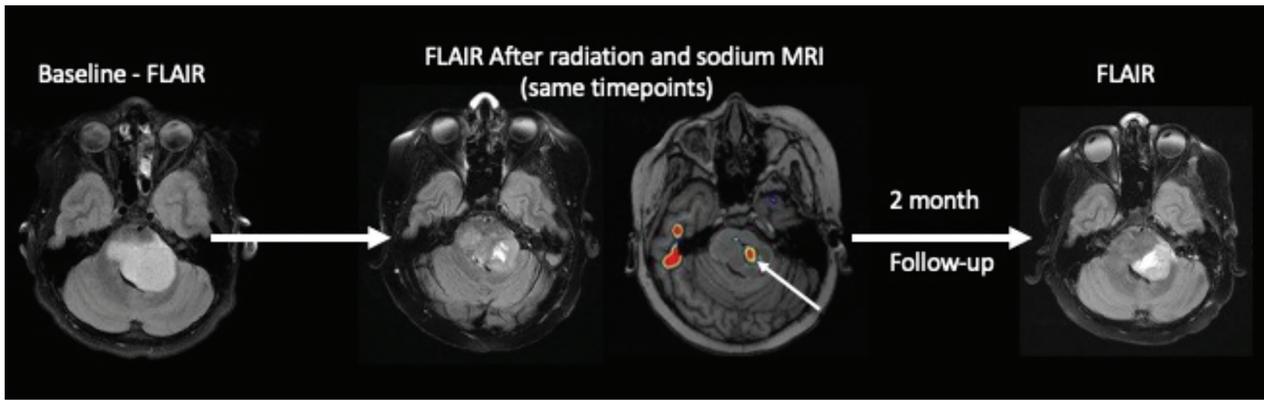
**FIG 3.** Diffuse midline glioma centered in the pons in a pediatric patient. T1 MPRAGE with tumor (*solid arrow*) centered in the pons extending into the left middle cerebellar peduncle. Subtraction of the 2 echo times demonstrates a focal region of elevated sodium in the tumor (*arrow*). Susceptibility artifact displays as elevated signal in the region of the right temporal bone (*dashed arrow*), which could potentially be misinterpreted as elevated tissue sodium concentration (same patient as above).



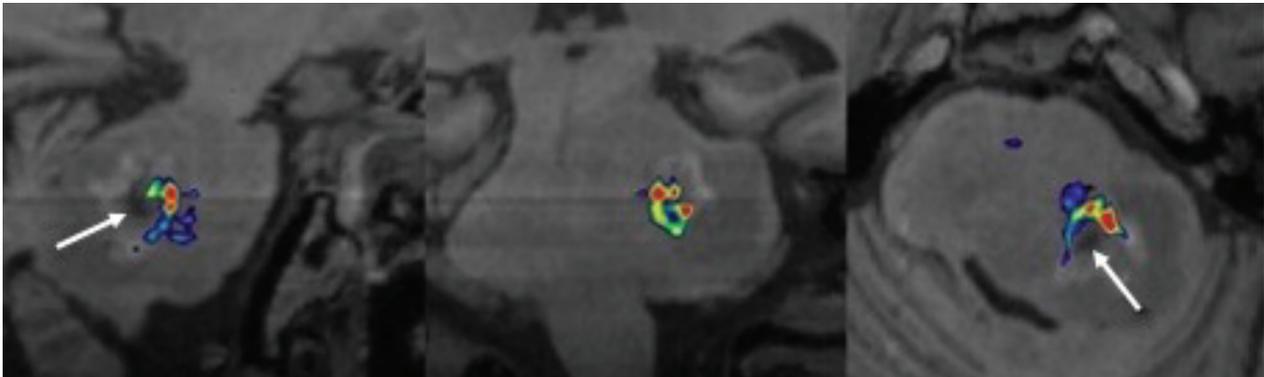
**FIG 4.** Thresholding the sodium concentrations in the pediatric patient above. Focal region of elevated sodium in the tumor (*arrow*). Thresholding allows precise depicting of the tumor.

anatomic prior insights from  $^1\text{H}$ -MRI allows anatomically-guided reconstruction, for example.<sup>32</sup> The TPI sequence that is relatively widely used offers  $\sim 3.4$ -mm isotropic resolution, somewhat

comparable with fMRI and, though justifiable in terms of its physiologic value, benefits from such anatomically-enhanced reconstruction.



**FIG 5.** Pediatric patients with diffuse midline glioma from above. The increased focus of sodium concentrations (*arrow*) in the tumor was acquired after radiation treatment; this region may represent the following: radioresistant region of the tumor (A) and region of tumor progression (B). Follow-up conventional MRI a further 2 months after radiation cessation demonstrates tumor progression in the region of the prior sodium elevation. This anecdote supports the hypothesis that elevated sodium signal represents an early biomarker of tumor progression/recurrence.



**FIG 6.** Using the T1-weighted sequence to further suppress necrotic regions (*arrows*) within the diffuse midline glioma (same patient as above) in a composite  $^1\text{H}$ -MRI and  $^{23}\text{Na}$ -MRI integration.

**Possible Clinical Utility.** Sodium MRI has the potential to be a more physiologically-specific imaging marker, compared with conventional  $^1\text{H}$ -MRI in determining treatment responses (Fig 5). Sodium MRI is sensitive to tumor recurrence that is not visualized on conventional  $^1\text{H}$ -MRI.<sup>33</sup> Thus,  $^{23}\text{Na}$ -MRI provides an objective imaging biomarker of tumor biologic activity.  $^{23}\text{Na}$  promises to offer a pivotal imaging finding in pediatric patients with brain tumors to allow neuro-oncologists the clear indication needed to modify treatment regimens, by providing a more definitive and physiologically-interpretable imaging marker of brain tumor progression and treatment efficacy.

**Future Directions.** High-field (7T) MRI allows technical advances in sodium imaging in the pediatric brain. The recent FDA-approved 7T Terra X magnet (Siemens) is equipped for heteronuclear imaging with a dual-frequency  $^{23}\text{Na}/^1\text{H}$  coil. 7T has advantages in sodium MRI such as decreasing image-acquisition time, improving SNR (approximately doubling that obtained at 3T), and thus increased resolution, though noting that increasing spatial resolution by a factor of 2 results in a decreased SNR by a factor of 8 if using a 3D isovolumetric sequence.<sup>23,34–37</sup> Thus, the potential 7T advantages must be tailored for the specific clinical question, such as in treatment of brain tumors, in which early

treatment effects can be better evaluated with  $^{23}\text{Na}$ -MRI due to improved SNR.

#### Multimodal T1-Weighted Enhancement of Sodium Images.

Figure 6 illustrates an interesting potential integration of  $^1\text{H}$ - and  $^{23}\text{Na}$ -MRI. In this example, the  $^1\text{H}$ -MRI (which is weighted by tissue T1 values) is used not only for image overlay of the  $^{23}\text{Na}$ -MRI but also as a weighted-mask, in which the T1-weighted  $^1\text{H}$ -MRI and the subtracted tissue sodium image are multiplied to yield a new sodium map, in which free fluid (and, in particular, necrotic regions) is suppressed by the combined effects of T1-weighted hyperintensity as well as reduced surviving signal on the dual-echo  $^{23}\text{Na}$ -MRI-weighted subtraction image. In this example, peripheral sodium signal encircling the necrotic core is clearly depicted. The eventual clinical utility of such approaches remains to be explored.

Limitations of the  $^{23}\text{Na}$ -MRI approach for estimating sodium concentrations in the pediatric brain include, but are not limited to, the time of imaging, with each of the dual TE sodium sequences needing  $\sim 10$  minutes, leading to a total acquisition time of  $\sim 20$  minutes, rendering data collection in children somewhat challenging without sedation. Acceleration techniques and yet-shorter TE acquisitions offer some promise for development.

Generally, the much lower sodium concentrations in the brain compared with  $^1\text{H}$  nuclei limit the use to at least 3T, with further studies needed to optimize (and thus realize the theoretic benefits of) implementation at higher field strengths, such as 7T. Currently, the postprocessing methods discussed above make a valuable contribution toward suppression of free fluid; however, complete specificity for intracellular-versus-extracellular sodium warrants further development.

## CONCLUSIONS

The technological advancements with improved pulse sequences and reconstruction methods allow combatting the challenges of measuring sodium concentrations in the pediatric brain. The sodium signal is a naturally inherent signal that does not require exogenous tracers or ionizing radiation, and in pediatric brain tumors, it is reflective of the cellular physiology of neoplastic cells in the tumor microenvironment. Here we present detailed imaging approaches to quantifying sodium concentrations in the pediatric brain that can be assessed in various CNS pathologies, with the focus on brain tumors. Future applications of  $^{23}\text{Na}$ -MRI include biopsy guidance with thresholding of the sodium concentrations to visualize potential regions of tumor with the highest mitotic activity. The major advantage of using  $^{23}\text{Na}$ -MRI will likely be to answer the complex questions on  $^1\text{H}$ -MRI that arise in monitoring treatment of pediatric patients with brain tumors: Is there residual or recurrent tumor, differentiating tumor progression versus treatment-related changes (pseudoprogression), and, in some cases at presentation, differentiating between a neoplastic process versus another mimicking pathology such as demyelinating disease.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

- Bottomley PA. Sodium MRI in man: technique and findings. In: Harris RK, ed. *Encyclopedia of Magnetic Resonance*. John Wiley & Sons, Ltd; 2007
- Gast LV, Platt T, Nagel AM, et al. Recent technical developments and clinical research-applications of sodium ( $^{23}\text{Na}$ ) MRI. *Prog Nucl Magn Reson Spectrosc* 2023;138-139:1-51 CrossRef Medline
- Deen SS, Riemer F, McLean MA, et al. Sodium MRI with 3D-cones as a measure of tumour cellularity in high grade serous ovarian cancer. *Eur J Radiol Open* 2019;6:156-62 CrossRef Medline
- Gerhalter T, Gast LV, Marty B, et al.  $^{23}\text{Na}$  MRI depicts early changes in ion homeostasis in skeletal muscle tissue of patients with Duchenne muscular dystrophy. *J Magn Reson Imaging* 2019;50:1103-13 CrossRef Medline
- Schepkin VD. Sodium MRI of glioma in animal models at ultrahigh magnetic fields. *NMR Biomed* 2016;29:175-86 CrossRef Medline
- LaVerde G, Nemoto E, Jungreis CA, et al. Serial triple quantum sodium MRI during non-human primate focal brain ischemia. *Magn Reson Med* 2007;57:201-05 CrossRef Medline
- Huhn K, Engelhorn T, Linker RA, et al. Potential of sodium MRI as a biomarker for neurodegeneration and neuroinflammation in multiple sclerosis. *Front Neurol* 2019;10:84 CrossRef Medline
- Madelin G, Regatte RR. Biomedical applications of sodium MRI in vivo. *J Magn Reson Imaging* 2013;38:511-29 CrossRef Medline
- Grapperon AM, Ridley B, Verschuere A, et al. Quantitative brain sodium MRI depicts corticospinal impairment in amyotrophic lateral sclerosis. *Radiology* 2019;292:422-28 CrossRef Medline
- Boada FE, Qian Y, Nemoto E, et al. Sodium MRI and the assessment of irreversible tissue damage during hyper-acute stroke. *Transl Stroke Res* 2012;3:236-45 CrossRef Medline
- Regnery S, Behl NG, Platt T, et al. Ultra-high-field sodium MRI as biomarker for tumor extent, grade and IDH mutation status in glioma patients. *Neuroimage Clin* 2020;28:102427 CrossRef Medline
- Ouwerkerk R, Bleich KB, Gillen JS, et al. Tissue sodium concentration in human brain tumors as measured with  $^{23}\text{Na}$  MR imaging. *Radiology* 2003;227:529-37 CrossRef Medline
- Bhatia A, Lee VK, Qian Y, et al. Quantitative sodium ( $^{23}\text{Na}$ ) MRI in pediatric gliomas: initial experience. *Diagnostics (Basel)* 2022;12:12232 CrossRef Medline
- Thulborn KR, Davis D, Adams H, et al. Quantitative tissue sodium concentration mapping of the growth of focal cerebral tumors with sodium magnetic resonance imaging. *Magn Reson Med* 1999;41:351-59 CrossRef
- Guan G, Zhao M, Xu X, et al. Abnormal changes in voltage-gated sodium channels subtypes NaV1.1, NaV1.2, NaV1.3, NaV1.6 and CaM/CaMKII pathway in low-grade astrocytoma. *Neurosci Lett* 2018;674:148-55 CrossRef Medline
- Gierzyng A, Pszczolkowska D, Walentynowicz KA, et al. Immune microenvironment of gliomas. *Lab Invest* 2017;97:498-518 CrossRef Medline
- Liberti MV, Locasale JW. The Warburg effect: how does it benefit cancer cells? *Trends Biochem Sci* 2016;41:211-18 CrossRef Medline
- Scharping NE, Delgoffe GM. Tumor microenvironment metabolism: a new checkpoint for anti-tumor immunity. *Vaccines (Basel)* 2016;4:46 CrossRef Medline
- Cong D, Zhu W, Kuo JS, et al. Ion transporters in brain tumors. *Curr Med Chem* 2015;22:1171-81 CrossRef Medline
- Rotoli D, Cejas MM, Maeso MD, et al. The Na, K-ATPase  $\beta$ -subunit isoforms expression in glioblastoma multiforme: moonlighting roles. *Int J Mol Sci* 2017;18:2369 CrossRef Medline
- Biller A, Badde S, Nagel A, et al. Improved brain tumor classification by sodium MR imaging: prediction of IDH mutation status and tumor progression. *AJNR Am J Neuroradiol* 2016;37:66-73 CrossRef Medline
- Schepkin VD, Ross BD, Chenevert TC, et al. Sodium magnetic resonance imaging of chemotherapeutic response in a rat glioma. *Magn Reson Med* 2005;53:85-92 CrossRef Medline
- Wilferth T, Gast L, Lachner S, et al. X-Nuclei MRI on a 7T MAGNETOM Terra: initial experiences. *MAGNETOM Flash* 2020;76
- Du J, Bydder M, Takahashi AM, et al. Short T2 contrast with three-dimensional ultrashort echo time imaging. *Magn Reson Imaging* 2011;29:470-82 CrossRef Medline
- Bydder M, Ali F, Saucedo A, et al. A study of 3D radial density adapted trajectories for sodium imaging. *Magn Reson Imaging* 2021;83:89-95 CrossRef Medline
- Feng L. Golden-angle radial MRI: basics, advances, and applications. *J Magn Reson Imaging* 2022;56:45-62 CrossRef Medline
- Qian Y, Panigrahy A, Laymon CM, et al. Short-T2 imaging for quantifying concentration of sodium ( $^{23}\text{Na}$ ) of bi-exponential T2 relaxation. *Magn Reson Med* 2015;74:162-74 CrossRef Medline
- Wilferth T, Gast LV, Stobbe RW, et al.  $^{23}\text{Na}$  MRI of human skeletal muscle using long inversion recovery pulses. *Magn Reson Imaging* 2019;63:280-90 CrossRef Medline
- Stobbe R, Boyd A, Smyth P, et al. Sodium intensity changes differ between relaxation- and density-weighted MRI in multiple sclerosis. *Front Neurol* 2021;12:693447 CrossRef Medline
- Blunck Y, Josan S, Taqdees SW, et al. 3D-multi-echo radial imaging of 23 Na (3D-MERINA) for time-efficient multi-parameter tissue compartment mapping. *Magn Reson Med* 2018;79:1950-61 CrossRef Medline
- Blunck Y, Kolbe SC, Moffat BA, et al. Compressed sensing effects on quantitative analysis of undersampled human brain sodium MRI. *Magn Reson Med* 2020;83:1025-33 CrossRef Medline
- Schramm G, Filipovic M, Qian Y, et al. Resolution enhancement, noise suppression, and joint T2\* decay estimation in dual-echo

- sodium-23 MR imaging using anatomically guided reconstruction.** *Magn Reson Med* 2024;91:1404–18 CrossRef Medline
33. De Leon-Benedetti L, Narayanan S, Lee VK, et al. **The use of sodium MRI in the diagnosis of an anaplastic astrocytoma during immunotherapy: a case report.** *Childs Nerv Syst* 2024;40:965–67 CrossRef Medline
34. Pahor M, Manini MC. **High resolution sodium imaging of human brain at 7T.** *Bone* 2008;23:1–7
35. Huang L, Bai J, Zong R, et al. **Sodium MRI at 7T for early response evaluation of intracranial tumors following stereotactic radiotherapy using the CyberKnife.** *AJNR Am J Neuroradiol* 2022;43:181–87 CrossRef Medline
36. Nunes Neto LP, Madelin G, Sood TP, et al. **Quantitative sodium imaging and gliomas: a feasibility study.** *Neuroradiology* 2018; 60:795–802 CrossRef Medline
37. Nagel AM, Bock M, Hartmann C, et al. **The potential of relaxation-weighted sodium magnetic resonance imaging as demonstrated on brain tumors.** *Invest Radiol* 2011;46:539–47 CrossRef Medline

# High-Resolution MR Imaging of the Parasellar Ligaments

Ian T. Mark, Myung-Ho In, Daehun Kang, Eric Stinson, John Huston, Yunhong Shu, Luciano Leonel, and Maria Peris Celda



## ABSTRACT

**SUMMARY:** The parasellar ligaments have been previously described in cadaver specimens and intraoperatively, but identification on MR imaging has eluded radiologists. Using high-resolution T2-weighted MR imaging, we identified the parasellar ligaments as T2-hypointense, bandlike structures that emanate from the medial wall of the cavernous sinus. Subsequent dissection of the same specimen provided matching anatomic images of the parasellar ligaments identified on MRI. This imaging finding is important because resection of the medial wall of the cavernous sinus has been tied to improved outcomes for gross total resection and endocrinologic remission of functioning pituitary adenomas.

**ABBREVIATIONS:** MWCS = medial wall of the cavernous sinus; NEX = number of excitations

The cavernous sinus is a complex structure that houses critical anatomy such as the ICAs and cranial nerves.<sup>1</sup> Recent studies have underscored the importance of the medial wall of the cavernous sinus (MWCS) in improving outcomes for functional pituitary adenoma surgery.<sup>2</sup> Within the cavernous sinus, the parasellar ligaments are important structures that extend from the MWCS to the ICAs and cranial nerves.<sup>3</sup> Despite their anatomic significance, the parasellar ligaments have remained elusive on MR imaging. The purpose of this report is to describe the imaging appearance of the parasellar ligaments on MR imaging.

### Imaging

All imaging was performed on a research-only compact 3T scanner (GE Healthcare).<sup>4</sup> A standard coronal T2-weighted sequence that mimics our clinically used protocol was initially run (slice thickness, 3 mm; TR, 3025 ms; TE, 101 ms; 160-mm field of view; number of excitations (NEX), 1; acquisition time 55 seconds). This clinical sequence demonstrated hazy T2-isointense signal in the cavernous sinus, immediately adjacent to the pituitary gland (Fig 1A). In our experience, ill-defined signal in the cavernous sinuses on the coronal T2-weighted sequence is a typical finding,

without having sufficient spatial resolution to identify the parasellar ligaments.

An additional high-resolution T2-weighted sequence was then performed to further resolve the structures within the cavernous sinus (slice thickness, 0.5 mm; TR, 4372 ms; TE, 100 ms; 160-mm field of view; NEX, 6; acquisition time, 21 minutes 13 seconds).

The higher-resolution imaging further resolved the ill-defined signal in the cavernous sinus as distinct T2-hypointense bands (Fig 1B). The bands run from the MWCS that directly abuts the lateral aspect of the pituitary gland and fan outward as they extend laterally toward the ICA and the lateral wall of the cavernous sinus. These distinct bands correlate with the parasellar ligaments (Fig 2), specifically the anterior horizontal ligament and periosteal ligament.

### Anatomy Report

For this study, we used 1 embalmed cadaveric specimen prepared as previously described.<sup>5</sup> Briefly, for the latex injection both the common carotid and vertebral arteries as well as internal jugular veins were isolated and cannulated with stainless steel cannulas and perfused with red and blue latex, respectively.

To expose and describe the pituitary region and ligaments, we performed a regular transsphenoidal approach. With a zero-degree endoscope, both middle and superior turbinates on both sides of the nasal cavity were lateralized, revealing the ostium of the sphenoid sinus. A Kerrison rongeur was used to enlarge the ostium and enter the sphenoid sinus cavity. Within the sphenoid sinus, the septum was drilled, and the bony landmarks, recesses, and prominences were exposed. The dissection proceeded by drilling the sellar prominence and moving laterally to the carotid prominence. Once the dura covering the pituitary gland and

Received November 19, 2024; accepted after revision January 11, 2025.

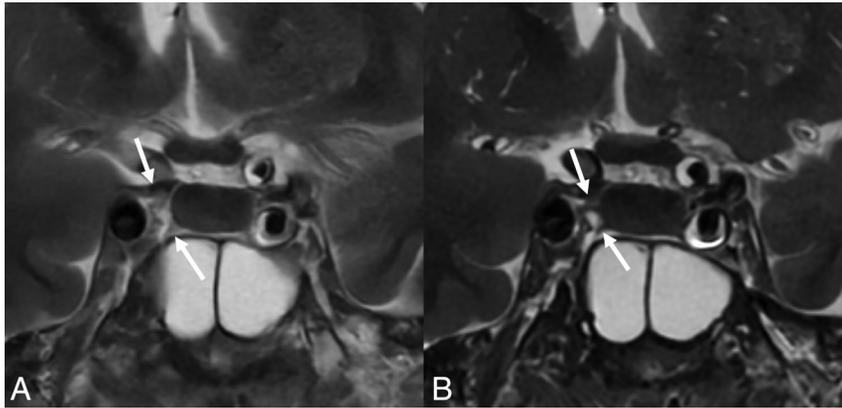
From the Departments of Radiology (I.T.M., M.I., D.K., E.S., J.H., Y.S.), Neurosurgery (L.L., M.P.C.), and Anatomy (L.L., M.P.C.), Mayo Clinic, Rochester, Minnesota.

Please address correspondence to Ian T. Mark, MD, Department of Radiology, Mayo Building, 2-50W, Rochester, MN 55905; e-mail: Mark.Ian@mayo.edu; @iantmark; @mayoradiology

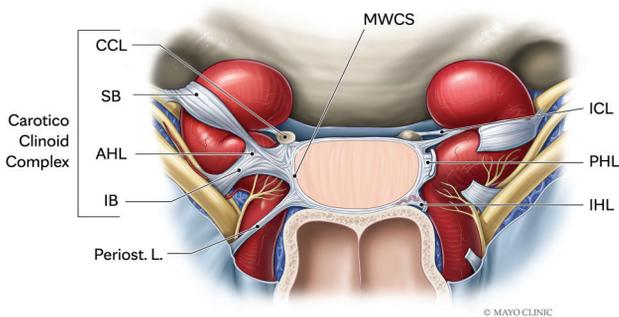
Indicates open access to non-subscribers at www.ajnr.org

Indicates article with supplemental data.

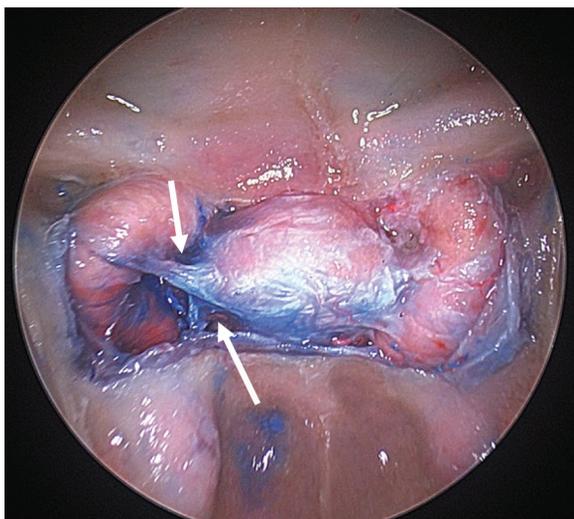
<http://dx.doi.org/10.3174/ajnr.A8658>



**FIG 1.** A, A standard 3-mm-thick coronal T2-weighted sequence shows hazy signal (arrows) in the cavernous sinus. B, High-resolution 0.5-mm slice thickness imaging resolves these hazy structures as distinct T2-hypointense bands that represent the parasellar ligaments (upper arrow indicates the anterior horizontal ligament; lower arrow, the periosteal ligament).



**FIG 2.** Illustration of the parasellar ligaments demonstrate the “candy wrapper” configuration that originates on the MWCS and fans out laterally as the ligaments extend to the cavernous segment of the ICA and the lateral wall of the cavernous sinus. CCL indicates carotid-clinoid ligament; AHL, anterior horizontal ligament and the associated SB, superior branch; IB, inferior branch; Periost. L., periosteal ligament; ICL, interclinoid ligament; PHL, posterior horizontal ligament; IHL, interclinoid ligament. This illustration has been modified from the original version.<sup>3</sup>



**FIG 3.** Endoscopic image from the same cadaver specimen from Fig 1. The distinct T2-hypointense bands in the right cavernous sinus seen on MRI correspond to the right parasellar ligaments. Specifically, the anterior horizontal ligament (upper arrow) and the periosteal ligament (lower arrow).

anterior bend of the ICA were exposed, a retractable knife was used to incise and expose the dura. The incision was performed at the level of the anterior intercavernous sinus, moving laterally to open the anterior wall of the cavernous sinus and its contents. Last, using 0° and 30° scopes, we dissected the pituitary region and cavernous sinus and photo documented them, preserving the pituitary ligaments and their landmarks, in particular their origins and insertions in the surrounding structures (Fig 3, Supplemental Data).

## DISCUSSION

The cavernous sinus is an important anatomic region to understand during resection of functioning pituitary adenomas. Resection of the MWCS, even in the absence of macroscopic invasion, has been shown to be associated with improved outcomes including gross total resection and endocrinologic remission.<sup>2</sup> Surgical resection of the MWCS is not without risk, with many important structures including the ICA and cranial nerves residing in the cavernous sinus. The parasellar ligaments have been described as key anatomic structures that connect the MWCS to the ICA and are therefore required to be transected during MWCS resection. Preoperative identification of the parasellar ligaments could therefore help with surgical planning to avoid carotid artery injury. While these structures have been described in cadaver specimens, we present a case of MR imaging of the parasellar ligaments with anatomic correlation.

The MWCS is a thin single-layer structure comprising the meningeal dura that is difficult to see on MRI.<sup>3</sup> The resection of the MWCS has been shown to be of increasing importance in achieving biochemical cure of functioning pituitary adenomas and to prevent recurrence in nonfunctioning adenomas.<sup>6-8</sup> Parasellar ligament transection is important in the resection of the MWCS from the surrounding neurovascular structures.<sup>3</sup>

Kehrli et al<sup>9</sup> theorized that the fetal pituitary gland is wrapped in a mesenchymal mass, of which the parasellar ligaments are remnants, a process that leads to variability in the ligaments. The parasellar ligaments are fan-shaped structures originating from the MWCS and extending to the tunica adventitia of the ICAs and laterally to the cranial nerves in what has been described as a “candy-wrapper” configuration.<sup>3</sup> The parasellar ligaments, when studied in cadaver models, have a variable configuration but are grossly organized into anterior and posterior groups. The anterior horizontal ligament and posterior horizontal ligament, identified on our specimen, reside in the anterior and posterior groups, respectively.

## Case Summary

We present high-resolution MR imaging of the parasellar ligaments. These ligaments are important structures that connect the MWCS to the ICAs and cranial nerves and have previously been described with gross visualization in cadaver models and

intraoperatively, but not previously, with MR imaging. The MWCS is an increasingly important structure in achieving endocrinologic remission following the resection of functioning pituitary adenomas; therefore, the parasellar ligaments have newly recognized relevance for surgical planning.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Truong HQ, Lieber S, Najera E, et al. **The medial wall of the cavernous sinus, Part 1: surgical anatomy, ligaments, and surgical technique for its mobilization and/or resection.** *J Neurosurg* 2019;131:122–30 CrossRef Medline
2. Pontes JP, Udoma-Udofa OC, de Oliveira JS, et al. **Efficacy and safety of cavernous sinus medial wall resection in pituitary adenoma surgery: a systematic review and a single-arm meta-analysis.** *Pituitary* 2023;26:340–51 CrossRef Medline
3. Serioli S, Plou P, Leonel LC, et al. **The “candy wrapper” of the pituitary gland: a road map to the parasellar ligaments and the medial wall of the cavernous sinus.** *Acta Neurochir (Wien)* 2023;165:3431–44 CrossRef Medline
4. Foo TK, Laskaris E, Vermilyea M, et al. **Lightweight, compact, and high-performance 3T MR system for imaging the brain and extremities.** *Magn Reson Med* 2018;80:2232–45 CrossRef Medline
5. Leonel LC, Carlstrom LP, Graffeo CS, et al. **Foundations of advanced neuroanatomy: technical guidelines for specimen preparation, dissection, and 3D-photodocumentation in a surgical anatomy laboratory.** *J Neurol Surg B Skull Base* 2021;82:e248–58 CrossRef Medline
6. de Macêdo Filho LJ, Diógenes AV, Barreto EG, et al. **Endoscopic endonasal resection of the medial wall of the cavernous sinus and its impact on outcomes of pituitary surgery: a systematic review and meta-analysis.** *Brain Sci* 2022;12:1354 CrossRef Medline
7. Cohen-Cohen S, Gardner PA, Alves-Belo JT, et al. **The medial wall of the cavernous sinus, Part 2: selective medial wall resection in 50 pituitary adenoma patients.** *J Neurosurg* 2019;131:131–40 CrossRef Medline
8. Borghei-Razavi H, Muhsen BA, Joshi K, et al. **Endoscopic extracapsular resection of an adrenocorticotrophic hormone-secreting macroadenoma with selective resection of the medial cavernous sinus wall.** *World Neurosurg* 2020;144:199 CrossRef Medline
9. Kehrlı P, Ali M, Reis M, et al. **Anatomy and embryology of the lateral sellar compartment (cavernous sinus) medial wall.** *Neurol Res* 1998;20:585–92 CrossRef Medline

# Cracking the Code of Calcification: How Presence and Burden among Intracranial Arteries Influence Stroke Incidence and Recurrence

Matteo Conte, Mohammed O. Alalfi, Riccardo Cau, Roberta Scicolone, Seemant Chaturvedi, Renu Virmani, Gianluca De Rubeis, Daniel Bos, and Luca Saba



## ABSTRACT

**BACKGROUND:** Intracranial atherosclerosis accounts for approximately 8% of all strokes in Western societies but the influence of arterial calcification on plaque instability is a topic of ongoing debate.

**PURPOSE:** Our purpose is to explore the association between the presence and burden of calcium in atherosclerotic plaques among intracranial arteries with the risk of clinical or silent stroke events through a systematic review and meta-analysis.

**DATA SOURCES:** Studies from PubMed and Embase investigating intracranial vessel calcification and stroke events were reviewed from inception through May 2024, adhering to PRISMA guidelines.

**STUDY SELECTION:** Eight longitudinal studies involving 7297 adult patients undergoing CT or CTA scans for symptomatic or asymptomatic intracranial atherosclerosis were included.

**DATA ANALYSIS:** Pooled odds ratios were calculated to assess the relationship between stroke events and either the presence or burden of intracranial arterial calcification. Quality assessment was conducted using QUADAS-2; overall evidence was established using GRADE system. Meta-analysis was performed using random-effects models.

**DATA SYNTHESIS:** After adjusting for confounding factors, the presence of intracranial arterial calcification was significantly associated with stroke incidence or recurrence (OR = 1.54; 95% CI 1.06–2.24,  $P < .001$ ). The strength of this association was found to be similar (OR = 1.56; 95% CI 1.11–2.19,  $P < .001$ ). A positive correlation was also found for calcium burden (OR = 1.31; 95% CI, 1.17–1.46;  $P < .001$ ). Heterogeneity was moderate for calcium presence ( $Q = 13.16$  and  $9.19$ ;  $I^2 = 62\%$  and  $42.61\%$ , respectively); negligible for burden analysis ( $Q = 6.01$ ;  $I^2 = 0.01\%$ ).

**LIMITATIONS:** Despite strict inclusion criteria, heterogeneity and variability in calcium scoring methods across studies were observed. The lack of segment-specific analysis may have limited clinical interpretation.

**CONCLUSIONS:** This meta-analysis demonstrates a weak yet present association between intracranial arterial calcification and stroke events. However, given the high prevalence of calcification in the general population, its role for stroke prediction has limited evidence. Future studies may focus on specific arterial segments and emerging calcification patterns to improve predictive accuracy

**ABBREVIATIONS:** GRADE = Grading of Recommendations Assessment, Development and Evaluation; HR = hazard ratio; HU = Hounsfield Unit; IAC = intracranial arterial calcification; MeSH = Medical Subject Headings; QUADAS-2 = Quality Assessment of Diagnostic Accuracy Studies-2

Cardiovascular diseases stand out as the primary cause of mortality and morbidity worldwide. Atherosclerosis is a precursor of cardiovascular diseases, including myocardial infarction and stroke.<sup>1</sup> In the past, most research focused on the degree of extracranial carotid stenosis, which is still considered the primary

parameter for therapeutic decision-making.<sup>2</sup> Over the last 10 years, robust evidence has demonstrated that even extracranial carotid artery plaque structure plays a key role in the risk stratification for stroke events.<sup>3</sup>

Considering large artery atherosclerosis, approximately one-fifth of ischemic strokes are due to extracranial carotid artery pathology, whereas intracranial atherosclerosis accounts for approximately

Received November 4, 2024; accepted after revision January 11, 2025.

From the University of Padua (M.C.), Padua, Italy; School of Medicine and Surgery (M.O.A.), University of Bologna, Bologna, Italy; Department of Radiology (R.C., R.S., L.S.), Azienda Ospedaliero Universitaria (A.O.U.), Monserrato (Cagliari), Italy; Department of Neurology & Stroke Program (S.C.), University of Maryland School of Medicine, Baltimore City, Maryland; CV Path Institute (R.V.), Gaithersburg, Maryland; Department of Radiology (G.D.R.), San Camillo, Rome, Italy; Departments of Epidemiology (D.B.), and Radiology and Nuclear Medicine (D.B.), Erasmus MC, University Medical Center, Rotterdam, the Netherlands; and Mayo Clinic (L.S.), Rochester, Minnesota.

Please address correspondence to Luca Saba, Department of Radiology, Azienda Ospedaliero Universitaria (A.O.U.), di Cagliari–Polo di Monserrato s.s., 554 Monserrato (Cagliari) 09045, Italy; e-mail: lucasaba@unica.it

Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8668>

## SUMMARY

**PREVIOUS LITERATURE:** Clinical implications of calcification status among intracranial arterial segments are poorly understood. Current studies show conflicting findings, showcasing lack of evidence and different methods of investigations. Histopathologic evidence suggests calcification may predict stroke by contributing to arterial stiffening and plaque instability, although its exact role remains unclear.

**KEY FINDINGS:** This study indicates a weak association between intracranial calcification and stroke incidence/recurrence. Calcium presence (OR = 1.56) and burden (OR = 1.31) were diversely associated with stroke, although the different heterogeneity between the models could have impacted the actual difference. Among included studies, the strongest effects are seen in the intracranial ICA.

**KNOWLEDGE ADVANCEMENT:** With this study, widespread intracranial calcium assessment has proved of little value in the clinical setting. Nonetheless, our findings underscore the potential of segment-specific calcification assessment. Additionally, emerging evidence on imaging modalities exploring newer calcium patterns may offer valuable insights into plaque instability and stroke risk.

8% of all strokes in Western societies.<sup>1</sup> Intracranial atherosclerosis is a more common cause of stroke in African Americans, Hispanics, and Asians.<sup>4,5</sup> Several imaging features have been shown to play a role in determining extracranial carotid artery plaque instability<sup>6,7</sup> but it is currently still unclear whether calcifications are part of these studied features. Some authors argue that plaque calcification is a protective factor,<sup>8</sup> associated with decreased rate of stroke events while others demonstrated a significant correlation between calcifications and stroke.<sup>9</sup> Little or no information is available on the status of calcifications among intracranial plaques.

The aim of this systematic review and meta-analysis is to analyze published articles on intracranial arterial segments to understand the role of plaque calcification in stroke events.

## MATERIALS AND METHODS

This meta-analysis of the literature was conducted according to the Preferred Reporting Items for Systemic Reviews and Meta-Analyses guidelines<sup>10</sup> and registered on the PROSPERO database (CRD42023459750).

### Search Strategy

A comprehensive review was conducted by using 2 different databases: PubMed and EMBASE. Assessment of the publications until May 2024 was done by 2 authors (M.C. and M.O.A.) and 3 different searches were performed on each database. The Medical Subject Heading (MeSH) terms used for the first search were: (vascular calcification) AND (pattern OR patterns) AND (carotid OR intracranial) AND (artery OR arteries) AND (atherosclerosis). The second search was conducted by using the following MeSH terms: (vascular calcification) AND (anterior cerebral OR middle cerebral OR posterior cerebral OR basilar; artery OR arteries) AND (atherosclerosis). The third search used the following MeSH terms: (cerebrovascular disease) AND (stroke) AND (vascular calcification). PubMed search yielded a total of 718 studies (96 for the first MeSH; 93 for the second; 529 for the third) while EMBASE search yielded 353 studies (304 for the first MeSH; 18 for the second; 31 for the third). A total of 302 Medline records and conference abstracts were then removed from EMBASE. No additional restrictions were then applied. In

addition to records identified through the search strategy, to ensure comprehensiveness we included 2 studies identified through expert recommendations.

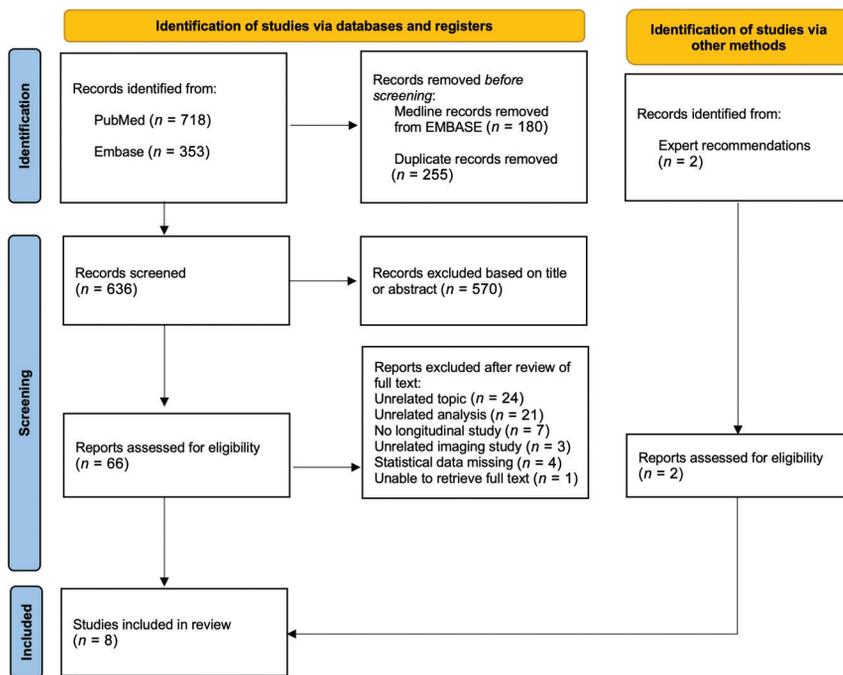
### Study Selection

After duplicates and not pertaining studies removal, 2 authors (M.C. and M.O.A.) independently screened the titles and abstracts based on the presence of search terms. Any conflict between the authors was resolved by the senior author (L.S.). The inclusion criteria were: adults undergoing CT scan/CTA of the brain for suspected stroke or adults undergoing CT scan/CTA of the brain for the purpose of longitudinal studies. Our goal was to target general population studies. To be ultimately included, both these groups needed to be assessed for the presence and/or the burden of intracranial vessel calcification and to be followed over time for clinical or silent stroke incidence or recurrence. Exclusion criteria were: articles not in English; review articles; case reports or case studies; studies with study population <10 patients; studies aiming to focus on specific target populations only (diabetics, chronic kidney disease); pediatric-population studies; studies with different imaging modalities than CT/CTA scans. Full texts were retrieved for studies that were eligible or possibly eligible. Studies that contained analytical data were evaluated. ORs or hazard ratios (HRs) from multivariate analyses were sought from original studies and used to discriminate the actual impact of calcification after the exclusion of potential confounders. Stroke events considered were either: clinical or silent, incidence, or recurrence. Ultimately, 7 prospective and 1 retrospective studies were investigated further (Fig 1).

### Quality Assessment

The presence of relevant biases in the included studies was evaluated according to Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) criteria.<sup>11</sup> The performed quality assessment included 4 domains: 1) patient selection; 2) index test; 3) reference standard; and 4) flow and timing. We used the proposed signaling questions to determine the risk of bias and concerns regarding the applicability, grading them as low, high or unclear in case of insufficient data (Fig 2).

We assessed the quality of the body of retrieved evidence by using the Grading of Recommendations Assessment, Development



**FIG 1.** Preferred Reporting Items for Systemic Reviews and Meta-Analyses flowchart of the literature search and study selection.

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Bos (2011)	😊	😊	😊	😊	😊	😊	😊
Gurel (2023)	😊	😊	😊	😊	😞	😊	😊
Lee (2014)	😊	😊	😊	😊	😞	😊	😊
Bugnicourt (2011)	😊	😊	😊	😊	😞	😊	😊
Kong (2019)	😊	😊	😊	😊	😞	😊	😊
Bos (2014)	😊	😊	😊	😊	😊	😊	😊
Quiney (2017)	😊	😊	😊	😊	😞	😊	😊
Wu (2020)	😊	😊	😊	😊	😞	😊	😊

😊 Low Risk    😞 High Risk    ? Unclear Risk

**FIG 2.** Summary of QUADAS-2.

and Evaluation (GRADE) methodologic system.<sup>12</sup> Quality of evidence was assessed based on the following domains: risk of bias, inconsistency of results, indirectness of evidence, imprecision of results, and publication bias. Because our review included only observational studies, quality of evidence was initially rated as “low quality.” Because our study had a low risk of bias due to confounding or selection bias, no downgrading was implemented. Because our study had small effect size and no substantial dose-response mechanism was recorded, no upgrading was implemented.

### Population Characteristics

The study included a total of 7297 patients from longitudinal studies, with an average of 936 patients per study. Of the recruited population, 3697 patients were randomly selected from prospective studies on asymptomatic patients while the remaining 3600 were selected after the initial presentation of either TIA or clinical stroke

events and followed either prospectively or retrospectively. The follow-up period ranged from 8 months to 10 years.

The population was further categorized based on the presence and eventual burden of calcification among intracranial arterial segments. For calcium presence, target population was compared with absence of calcium based on individual Hounsfield unit (HU) cut-off as reported in the Table. For calcium burden, higher calcium volume (crude HU, upper tertile, upper quartile) or severe intracranial calcification (IAC) scores were compared with lower calcium volume or mild IAC scores as reported in the Supplemental Data.

### Statistical Analysis

This study aimed to compare 2 distinctive aspects: 1) association between presence of calcium and incidence or recurrence of stroke and 2) association between burden of calcium and incidence or recurrence of stroke.

The study was initially conducted by using Jamovi statistical software (Version 1.6.23.0) with the MAJOR (Meta-Analysis JAMOVI) module. A refined version was obtained by using the R statistical software (Version 4.4.2). Meta-analyses were performed with the metafor package, and forest plots were generated by using the forestplot package. The first model was realized with unadjusted raw patient data grouping total vascular events in calcified vessels (group A) out of total calcified vessels (group B), and total vascular events in control vessels (group C) out of total control vessels (group D). Effect size models

were realized converting adjusted ORs or HRs from original articles in  $\ln(\text{OR})$  and back-converting the data in ORs for clearer interpretation. We investigated included studies to identify any discrepancies between reported ORs/HRs and expected values, based on study protocols or preregistered information when available. Variance was based on sample sizes from original articles. A random-effects model was used to account for heterogeneity across studies, under the assumption that the true effect size varied between studies. The model was fitted by using the restricted maximum likelihood estimation method, implemented in the metafor package. This iterative approach reconstructs the model parameters by optimizing the log-likelihood function, providing unbiased estimates of both the average effect size and the between-study variance ( $\tau^2$ ). Heterogeneity among studies was assessed by using the Cochran Q statistic,  $I^2$  statistic,  $I^2$  statistic, and funnel plots. The Q statistic tests the null hypothesis that effect sizes are equal across studies, with  $Q > 14.07$

### Association between calcium presence and subclinical/clinical events across intracranial arterial segments

Author	Segments Studied	Calcium Evaluation	Population <sup>a</sup>	Association	Effect Estimate	P Value
Bos, 2011 <sup>13</sup>	1770 (ICA only)	Cutoff: 130 HU	Asymptomatic	Subclinical stroke incidence <sup>b</sup>	OR 1.44 (0.59, 3.48)	<.01
Gurel, 2023 <sup>16</sup>	9055	Cutoff: 130 HU	Cohort #1: Patient with IS/TIA	Clinical stroke recurrence (FU >10 years)	OR 1.10 (0.80–1.53)	<.05
		Cutoff: 130 HU	Cohort #2: Asymptomatic patients	Clinical stroke incidence (FU >10 years)		
		Cutoff: 130 HU	Combined cohorts	Clinical stroke incidence/recurrence (FU >10 years)		
Lee, 2014 <sup>17</sup>	11,240	Cutoff: 90 HU	IS/TIA	Clinical stroke recurrence within <2 weeks	OR 1.49 (0.83–2.67)	.029
Bos, 2014 <sup>14</sup>	4646 (ICA only)	Cutoff 130 HU	Asymptomatic	Clinical or silent stroke incidence (FU 6 years)	OR 4.25 (1.55–11.66)	<.05
Quiney, 2017 <sup>19</sup>	1287	Cutoff 130 HU	IS/TIA	Clinical downstream stroke incidence	OR 2.2 (1.2–3.9)	.009
Wu, 2020 <sup>20</sup>	7634	Cutoff 130 HU	IS/TIA	Clinical stroke recurrence	OR 1.23 (0.57–2.66)	.599

Note:—FU indicates follow-up; IS/TIA, ischemic stroke or transient ischemic attack.

<sup>a</sup>Patients with absent intracranial calcifications constituted control group for all included studies.

<sup>b</sup>Presence of cerebral infarcts after 8 months.

indicating relevant heterogeneity. Funnel plots were derived from effect size (x-axis) and standard error (y-axis). A symmetric funnel plot indicates no publication bias; asymmetrical funnel plot indicates the possibility of publication bias. All data used in this systematic review and meta-analysis were extracted from publicly available material, ensuring transparency and reproducibility of the findings.

## RESULTS

The search identified 1071 studies for inclusion in the methodologic review. After duplicates and not pertaining studies were removed, 636 studies were screened based on set inclusion and exclusion criteria, resulting in the inclusion of 8 studies<sup>13–20</sup> (Fig 1). Only CT/CTA-based prospective studies were included. A total of 7297 patients were either randomly selected from large cohort studies or selected for follow-up after the initial presentation of strokelike symptoms. The average weighted age for the included population was 66.3 years old, with balanced weighted percentage between men and women (47.1% men, 52.9% women). Risk factors for cardiovascular diseases across the entire population were weighted with study size and distributed as follows: hypertension (70.01%); hypercholesterolemia (55.25%); type 2 diabetes (20.45%); positive smoking or current status (38.06%); chronic kidney disease (8.37%).

Patients were categorized according to the presence of calcium and the burden of calcium distribution. Except for 2 studies that focused on intracranial ICAs, the results section is based on calcium distribution across: bilateral intracranial carotid arteries, middle cerebral arteries, posterior cerebral arteries, and vertebro-basilar systems. Stroke incidence was the main target of 4 cohorts composed of stroke-free patients ( $n = 3697$  asymptomatic;  $n = 409$  initial stroke suspicion, ultimately categorized as stroke-free); stroke recurrence was the main target of 4 cohorts composed of patients with confirmed stroke or TIA ( $n = 3600$ ).

### Quality Assessment and Heterogeneity

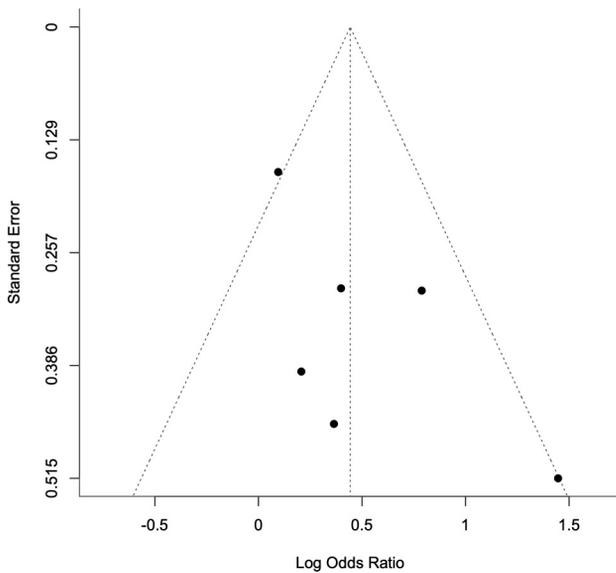
A summary of the quality, according to QUADAS-2 criteria assessment for the included studies, is summarized in Fig 2. The risk of bias for patient selection was low in all studies ( $n = 8$ ) due

to a consecutive sample of patients and a detailed description of patient enrollment. High risk of applicability concerns was detected in 6 studies due to selection of a specific target population limiting the generalizability of the findings. The risk of bias for index test and applicability concerns for the index test were found to be low in all studies ( $n = 8$ ). The risk of bias for the reference standard and applicability concerns for the reference standard were found to be low in all studies ( $n = 8$ ). Low risk of bias for flow and timing was found in all studies ( $n = 8$ ).

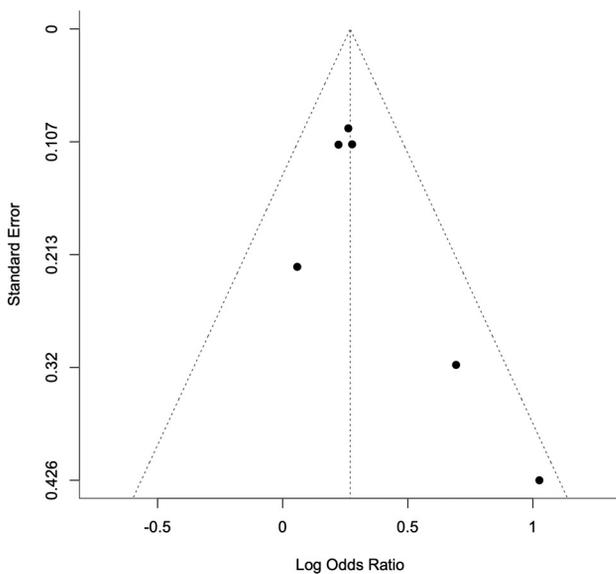
Overall evidence was qualified by using GRADE for observational studies. Risk of bias was low in all studies ( $n = 8$ ). Inconsistency was low in 6 studies; 2 studies proved greater association between calcium and stroke events, highlighting key differences in study populations and/or target of investigations. Indirectness was low in all studies ( $n = 8$ ). Imprecision was low in 4 studies that showed a narrow CI ( $<1$ ); moderate in 4 studies ( $CI >1$ ). Publication bias was low in all studies ( $n = 8$ ).

For calcium presence, 2 separate models were made from 6 of the 8 included studies. The first model evaluated the potential association between calcium and stroke events and contained nonadjusted data. The model had Q value of 13.16 ( $P = .02$ ), indicating that heterogeneity exists across the included studies. A higher proportion of total variation across studies was due to true difference between studies rather than chance alone ( $I^2 62%$ ). Specifically, the total variation results in triple compared with what would be expected from random error ( $H^2 2.63$ ). The second model evaluated the strength of the association and contained adjusted data when possible. The model had a Q value of 9.19 ( $P = .10$ ), with some variation expected to be true ( $I^2 42.61%$ ). Accordingly, total variation results were the same as to what would be expected from random error ( $H^2 1.74$ ). Funnel plot of the effect size model is reported on Fig 3.

For calcium burden, a single effect size model was made from 6 of 8 included studies to establish association with either stroke incidence or recurrence. The model had a Q value of 6.01 ( $P = .30$ ), with no variation expected to be true ( $I^2 0.01%$ ). Accordingly, total variation results were the similar as to what would be expected from random error ( $H^2 1.00$ ). Funnel plot is reported on Fig 4.



**FIG 3.** Funnel plot of the effect size for calcium presence and stroke incidence or recurrence.

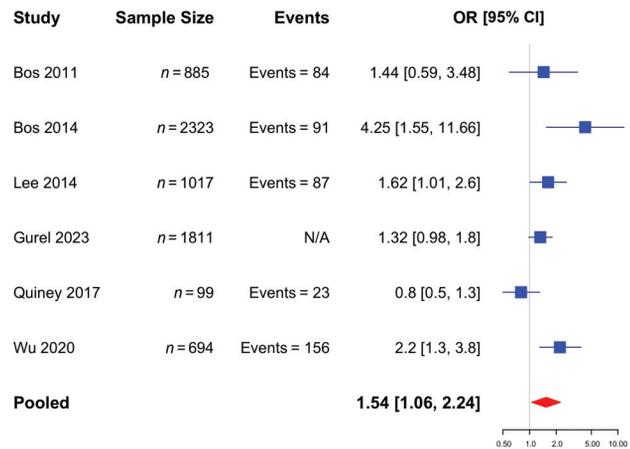


**FIG 4.** Funnel plot of the effect size for calcium burden and stroke incidence or recurrence.

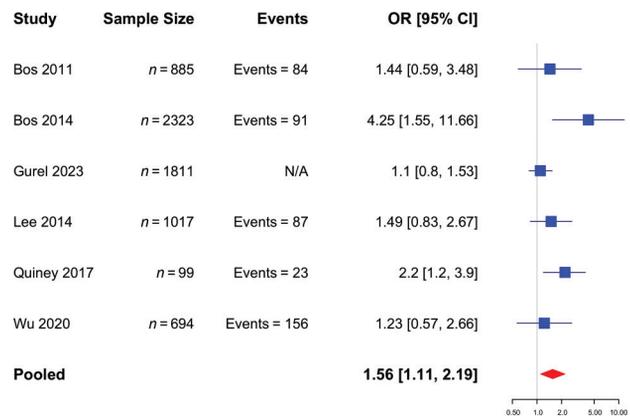
### Calcium Presence and Stroke Incidence or Recurrence

Individual evaluation for calcium presence, investigation methods, target and comparison group, association type (clinical or silent stroke), and effect estimate (OR, 95% CI, *P* value) are reported in the Table.

Of 8 included studies, 6 studies assessed calcium presence. Individual study characteristics are reported in the Table. A total of 6894 extracted patients were considered, with a total of 35,632 segments analyzed. Pooled data were used to determine the association between calcium presence and stroke incidence or recurrence within the selected population. We included both early (<2 weeks) and late recurrence. The follow-up time to monitor late recurrence ranged from 8 months to 10 years.



**FIG 5.** Forest plot of studies evaluating the association between calcium presence and stroke incidence or recurrence in intracranial arteries.

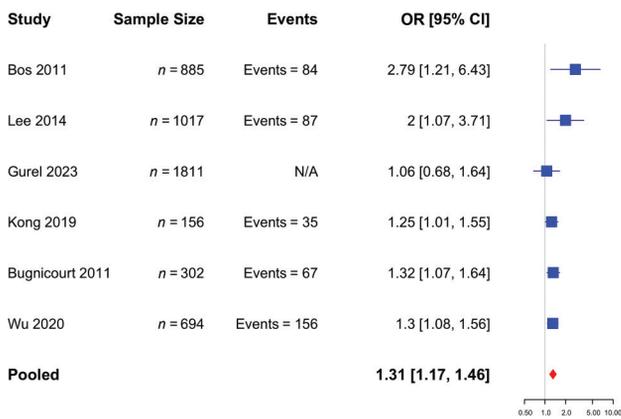


**FIG 6.** Forest plot of effect size of studies evaluating the association strength between calcium presence and stroke incidence or recurrence in intracranial arteries.

Two statistical models were created. The first model assessed the potential association between calcium and stroke events by using unadjusted data. After adjusting for notorious risk factors of plaque destabilization, multivariate ORs from included original studies were combined to generate a random-effect model. In the presence of calcium among intracranial atherosclerotic plaques, our model predicted a positive correlation with stroke incidence or recurrence (OR = 1.54, 95% CI 1.06–2.24) (*P* = .001). The relative forest plot is presented in Fig 5. Second, effect sizes were computed to establish the actual strength of the previous correlation. After adjusting for notorious risk factors of plaque destabilization, multivariate ORs from included original studies were transformed into log(OR) and combined to generate a random-effect model. In the presence of calcium among intracranial atherosclerotic plaques, our model predicted a positive correlation with stroke incidence or recurrence (OR = 1.56, 95% CI 1.11–2.19) (*P* ≤ .001). The relative forest plot is presented in Fig 6.

### Calcium Burden and Stroke Incidence or Recurrence

Individual evaluation for calcium burden, investigation methods, target and comparison group, association type (clinical or subclinical



**FIG 7.** Forest plot of effect size of studies evaluating the association strength between calcium burden and stroke incidence or recurrence in intracranial arteries.

stroke), and effect estimate (OR, 95% CI, *P* value) are reported in the Supplemental Data.

Of 8 included studies, 6 studies assessed calcium burden. Individual study characteristics are reported in the Supplemental Data. A total of 4885 extracted patients were considered, with a total of 32,905 segments analyzed. Pooled data were used to determine the association between calcium burden and stroke incidence or recurrence within the selected population. We included both early (<2 weeks) and late recurrence. The follow-up time to monitor late recurrence ranged from 8 months to 10 years.

Effect sizes were computed to establish the strength of correlation between higher concentrations of calcium and lower ones with stroke. After adjusting for notorious risk factors of plaque destabilization, multivariate ORs from included original studies were transformed into log(OR) and combined to generate a random-effect model. With higher calcium concentrations across intracranial plaques, our model predicted a positive correlation with stroke incidence or recurrence (OR=1.31, 95% CI, 1.17–1.46) (*P* ≤ .001). The relative forest plot is presented in Fig 7.

## DISCUSSION

In a meta-analysis of 8 studies involving 7297 patients, our study has demonstrated that while calcification is associated with stroke incidence or recurrence, the association, according to our effect size model, is weak. Furthermore, while our study showed an existing association between calcium burden and stroke incidence or recurrence, this association is also weak.

Several studies have been performed to discern the association between intracranial calcification and stroke. While some authors found a relevant association between stroke and intracranial calcification,<sup>21–24</sup> other authors believe that the association is not meaningful.<sup>25–27</sup>

From a histopathologic point of view,<sup>28–30</sup> intimal atherosclerotic calcification is the most common pathway of calcium accumulation. After diffuse thickening of the tunica intima, vascular supply to the plaque becomes compromised, leading to the development of local foci of necrosis. These necrotic foci stimulate calcium deposition, while the release of proinflammatory cytokines

supports the conversion to a fibroinflammatory lipid plaque. This process is associated with plaque instability and an increased risk of rupture.<sup>30</sup> According to this pathologic process, calcium presence should be considered a potential predictor for stroke. However, evidence suggests that plaque calcification may manifest as an actively regulated process similar to bone formation, potentially initiated by calcifying muscular cells within the plaque.<sup>31</sup> Whether this process could stabilize the plaque is more challenging to predict.

In the light of these studies, our findings suggest a significant association between calcium and the incidence or recurrence of stroke events. After adjusting for potential confounders, our effect size analysis highlighted the strength of this association as the random effect remained consistent between the 2 models (from an OR of 1.54 in the first model to an OR of 1.56). Considering the moderate heterogeneity across included studies and the relative high prevalence of intracranial calcifications in the general population, these findings suggest a weak, yet present, association between calcification and stroke incidence or recurrence.

A more valuable approach appears to be focusing solely on the intracranial segment of the ICA. According to the study findings of Bos et al,<sup>13,14</sup> the strength of the association between stroke incidence and calcium presence among the intracranial ICAs was more consistent compared with other included studies that investigated the totality of intracranial segments. The odds of clinical or silent strokes in these groups increased from 1.44 (0.59–3.48) over an 8-month follow-up to 4.25 (1.55–11.66) over a 6-year follow-up.

The mechanism behind this relationship could be related to calcium-induced arterial stiffening. Stiffening of larger intracranial arteries like the ICA might result in cerebral hemodynamic impairment in the form of elevated generalized artery flow velocity elevation.<sup>32</sup> Abnormally elevated flow velocity could potentially destabilize plaque homeostasis, increasing shear stress and leading to rupture. This theory was investigated by Kamel et al,<sup>33</sup> who found a strong positive correlation between upstream large-artery atherosclerosis and downstream cerebral infarctions classified as cryptogenic according to current diagnostic criteria.

As previously theorized, calcification within intracranial vessels may be associated with arterial stiffening, hemodynamic impairment and plaque destabilization. Whether this association could become stronger when considering calcium burden rather than presence alone within intracranial arteries was our second aim for this meta-analysis. According to our random effect model, higher calcium burden influences the likelihood of stroke events by 1.31 times compared with lower calcium burden. Once more, the strongest effect comes from the evaluation of calcium burden in the intracranial segment of the ICA alone, showing that patients within the highest quartile of calcium volume (mm<sup>3</sup>) are 2.79 times more likely to develop a subclinical stroke within 8 months interscan period.

In summary, random effects models from both calcium presence (OR = 1.56) and burden (OR = 1.31) demonstrate a low, yet present, association with stroke events. Ultimately, evaluating burden of calcium across the intracranial segments failed to strengthen the previously established association, which seems to

decrease in strength. This finding could be explained by the different heterogeneity of the models rather than an actual difference (from a  $I^2$  of 42.61% in the first model to a  $I^2$  of 0.01%). Given the high prevalence of intracranial calcifications and overall low robustness of the present meta-analysis, our findings empower that widespread intracranial arterial calcification assessment may represent a weak indicator of stroke incidence or recurrence in clinical settings. Nonetheless, focusing on calcium presence on specific target vessels could be useful to predict strokes over time. To empower calcium assessment in clinical practice, more evidence from longitudinal studies is therefore needed.

### **Microcalcifications: The Definitive Response?**

Our meta-analysis focused on definite numerical cutoff to establish presence or absence of calcium. Perhaps size does not matter, and a different approach is needed. Particles of calcium that are normally not visible among standard imaging could give insight into stroke prediction and guide management in clinical practice. Numerous studies identify microcalcifications as a risk factor for coronary artery plaque instability.<sup>34-37</sup> Histopathologic studies on coronary arteries indicate that high risk lesions are nonobstructive plaques with softer, lipid-rich cores, and a thin fibrous cap containing microcalcifications. It is worth noting that even stable plaques may rupture in the presence of microcalcifications.<sup>37</sup> A recent pilot study on 18 patients with recurrent stroke or TIA on the carotid artery found a significant association between microcalcifications and ipsilateral stroke recurrence (adjusted for age OR = 1.24, 95% CI 1.03–1.50,  $P = <.10$ ).<sup>38</sup> Interestingly, the same study also reported a minor association between macrocalcifications burden and ipsilateral stroke recurrence (adjusted for age OR = 1.12, 95% CI 1.06–1.17  $P = <.10$ ).<sup>38</sup> Whether these entities could represent the definitive response for future stroke prediction among intracranial arteries is yet to be established.

### **Limitations**

Despite the insights drawn from our meta-analysis, several limitations should be addressed. First, even after strict exclusion criteria, the heterogeneity across studies persisted, underscoring the need for standardized methodologies and rigorous study designs in future research. Second, valuable data loss among excluded studies could have affected the findings of this meta-analysis. Our study was also limited by the lack of a common unified calcium scoring pattern. Last, analysis on individual arterial segments was not feasible due to lack of segment-specific data in the screened literature, undermining the precision and clinical meaningfulness of the present study. As such, to collect valuable data in a highly prevalent condition, future studies should be performed in longitudinal fashion and encourage the usage of a unified calcium scoring system. Strength of association investigation should be performed throughout subgroup analysis of individual arterial segments to enhance statistical power.

### **CONCLUSIONS**

The intricate relationship between calcium accumulation and atherosclerotic plaque development presents a multifaceted challenge in understanding its role in stroke incidence and recurrence. Our research underscores weak, yet present, association

between calcium and stroke events across intracranial arteries. Whether this association has an actual impact on clinical practice remains still unclear.

This meta-analysis reveals that widespread assessments of intracranial arterial calcification may offer limited value in stroke prediction, urging a more targeted approach. Notably, the focus on the intracranial segment of the ICA and the consideration of microcalcifications could provide clinically useful insights into plaque instability and stroke prediction and longitudinal analysis is needed.

### **ACKNOWLEDGMENTS**

The authors extend their sincere gratitude to everyone who contributed to the successful completion of this study. No financial support or specific funding was received for the conduct and preparation of this study.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

### **REFERENCES**

1. Huang Y, Li Y, Pan H, et al. **Global, regional, and national burden of neurological disorders in 204 countries and territories worldwide.** *J Glob Health* 2023;13:04160 CrossRef Medline
2. Barnett HJ, Taylor DW, Eliasziw M, et al. **Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators.** *N Engl J Med* 1998;339:1415–25 CrossRef Medline
3. Saba L, Saam T, Jager HR, et al. **Imaging biomarkers of vulnerable carotid plaques for stroke risk prediction and their potential clinical implications.** *Lancet Neurol* 2019;18:559–72 CrossRef Medline
4. White H, Boden-Albala B, Wang C, et al. **Ischemic stroke subtype incidence among whites, blacks, and Hispanics: the Northern Manhattan Study.** *Circulation* 2005;111:1327–31 CrossRef Medline
5. Wang Y, Zhao X, Liu L; CICAS Study Group, et al. **Prevalence and outcomes of symptomatic intracranial large artery stenoses and occlusions in China: the Chinese Intracranial Atherosclerosis (CICAS) Study.** *Stroke* 2014;45:663–69 CrossRef Medline
6. Yuan C, Zhang SX, Polissar NL, et al. **Identification of fibrous cap rupture with magnetic resonance imaging is highly associated with recent transient ischemic attack or stroke.** *Circulation* 2002;105:181–85 CrossRef Medline
7. Saba L, Nardi V, Cau R, et al. **Carotid artery plaque calcifications: lessons from histopathology to diagnostic imaging.** *Stroke* 2022;53:290–97 CrossRef Medline
8. Nandalur KR, Hardie AD, Raghavan P, et al. **Composition of the stable carotid plaque: insights from a multidetector computed tomography study of plaque volume.** *Stroke* 2007;38:935–40 CrossRef Medline
9. Saba L, Chen H, Cau R, et al. **Impact analysis of different CT configurations of carotid artery plaque calcifications on cerebrovascular events.** *AJNR Am J Neuroradiol* 2022;43:272–79 CrossRef Medline
10. Page MJ, McKenzie JE, Bossuyt PM, et al. **The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.** *BMJ* 2021;372:n71 CrossRef Medline
11. Whiting PF, Rutjes AW, Westwood ME, et al. **QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies.** *Ann Intern Med* 2011;155:529–36 CrossRef Medline
12. Berkman ND, Lohr KN, Ansari M, et al. **Grading the strength of a body of evidence when assessing health care interventions.** In: Santaguida PL, Raina P, McPheeters ML, et al., eds. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews.* Agency for

- Healthcare Research and Quality (US); 2013. Chapter 8. <https://www.ncbi.nlm.nih.gov/books/NBK174881/>.
13. Bos D, Ikram MA, Elias-Smale SE, et al. **Calcification in major vessel beds relates to vascular brain disease.** *Arterioscler Thromb Vasc Biol* 2011;31:2331–37 CrossRef Medline
  14. Bos D, Portegies ML, van der Lugt A, et al. **Intracranial carotid artery atherosclerosis and the risk of stroke in whites: the Rotterdam Study.** *JAMA Neurol* 2014;71:405–11 CrossRef Medline
  15. Bugnicourt JM, Leclercq C, Chillon JM, et al. **Presence of intracranial artery calcification is associated with mortality and vascular events in patients with ischemic stroke after hospital discharge: a cohort study.** *Stroke* 2011;42:3447–53 CrossRef Medline
  16. Gurel K, Khasiyev F, Spagnolo-Allende A, et al. **The role of intracranial artery calcification (IAC) in stroke subtype and risk of vascular events.** *J Stroke Cerebrovasc Dis* 2023;32:107185 CrossRef Medline
  17. Lee JG, Lee KB, Roh H, et al. **Intracranial arterial calcification can predict early vascular events after acute ischemic stroke.** *J Stroke Cerebrovasc Dis* 2014;23:e331–37 CrossRef Medline
  18. Kong WY, Tan BY, Ellis ES, et al. **Intracranial artery calcium burden predicts recurrent cerebrovascular events in transient ischaemic attack patients.** *J Stroke Cerebrovasc Dis* 2019;28:2332–36 CrossRef Medline
  19. Quiney B, Ying SM, Hippe DS, et al. **The association of intracranial vascular calcification and stenosis with acute ischemic cerebrovascular events.** *J Comput Assist Tomogr* 2017;41:849–53 CrossRef Medline
  20. Wu X, Bos D, Ren L, et al. **Intracranial arterial calcification relates to long-term risk of recurrent stroke and post-stroke mortality.** *Front Neurol* 2020;11:559158 CrossRef Medline
  21. Taoka T, Iwasaki S, Nakagawa H, et al. **Evaluation of arteriosclerotic changes in the intracranial carotid artery using the calcium score obtained on plain cranial computed tomography scan: correlation with angiographic changes and clinical outcome.** *J Comput Assist Tomogr* 2006;30:624–28 CrossRef Medline
  22. Sohn YH, Cheon HY, Jeon P, et al. **Clinical implication of cerebral artery calcification on brain CT.** *Cerebrovasc Dis* 2004;18:332–37 CrossRef Medline
  23. Power A, Chan K, Haydar A, et al. **Intracranial arterial calcification is highly prevalent in hemodialysis patients but does not associate with acute ischemic stroke.** *Hemodial Int* 2011;15:256–63 CrossRef Medline
  24. Erbay S, Han R, Baccei S, et al. **Intracranial carotid artery calcification on head CT and its association with ischemic changes on brain MRI in patients presenting with stroke-like symptoms: retrospective analysis.** *Neuroradiology* 2007;49:27–33 CrossRef Medline
  25. de Weert TT, Cakir H, Rozie S, et al. **Intracranial internal carotid artery calcifications: association with vascular risk factors and ischemic cerebrovascular disease.** *AJNR Am J Neuroradiol* 2009;30:177–84 CrossRef Medline
  26. Chen XY, Lam WW, Ng HK, et al. **Intracranial artery calcification: a newly identified risk factor of ischemic stroke.** *J Neuroimaging* 2007;17:300–03 CrossRef Medline
  27. Bos D, Arshi B, van den Bouwhuisen QJA, et al. **Atherosclerotic carotid plaque composition and incident stroke and coronary events.** *J Am Coll Cardiol* 2021;77:1426–35 CrossRef Medline
  28. Nakahara T, Dweck MR, Narula N, et al. **Coronary artery calcification: from mechanism to molecular imaging.** *JACC Cardiovasc Imaging* 2017;10:582–93 CrossRef Medline
  29. Otsuka F, Sakakura K, Yahagi K, et al. **Has our understanding of calcification in human coronary atherosclerosis progressed?** *Arterioscler Thromb Vasc Biol* 2014;34:724–36 CrossRef Medline
  30. Karwowski W, Naumnik B, Szczepański M, et al. **The mechanism of vascular calcification – a systematic review.** *Med Sci Monit* 2012;18:RA1–11 CrossRef Medline
  31. Jeziorska M, McCollum C, Wooley DE. **Observations on bone formation and remodelling in advanced atherosclerotic lesions of human carotid arteries.** *Virchows Arch* 1998;433:559–65 CrossRef Medline
  32. Wu X, Wang L, Zhong J, et al. **Impact of intracranial artery calcification on cerebral hemodynamic changes.** *Neuroradiology* 2018;69:357–63 CrossRef Medline
  33. Kamel H, Gialdini G, Baradaran H, et al. **Cryptogenic stroke and nonstenosing intracranial calcified atherosclerosis.** *J Stroke Cerebrovasc Dis* 2017;26:863–70 CrossRef Medline
  34. Joshi NV, Vesey AT, Williams MC, et al. **18F-fluoride positron emission tomography for identification of ruptured and high-risk coronary atherosclerotic plaques: a prospective clinical trial.** *Lancet* 2014;383:705–13 CrossRef Medline
  35. Hutcheson JD, Maldonado N, Aikawa E. **Small entities with large impact: microcalcifications and atherosclerotic plaque vulnerability.** *Curr Opin Lipidol* 2014;25:327–32 CrossRef Medline
  36. Chen W, Dilsizian V. **Targeted PET/CT imaging of vulnerable atherosclerotic plaques: microcalcification with sodium fluoride and inflammation with fluorodeoxyglucose.** *Curr Cardiol Rep* 2013;15:364 CrossRef Medline
  37. Corti A, De Paolis A, Grossman P, et al. **The effect of plaque morphology, material composition and microcalcifications on the risk of cap rupture: a structural analysis of vulnerable atherosclerotic plaques.** *Front Cardiovasc Med* 2022;9:1019917 CrossRef Medline
  38. Bhakta S, Tarkin JM, Chowdhury MM, et al. **Carotid atherosclerotic plaque microcalcification is independently associated with recurrent neurovascular events: a pilot study.** *Int J Stroke* 2024;19:1155–61 CrossRef Medline

# Location-Specific Net Water Uptake and Malignant Cerebral Edema in Acute Anterior Circulation Occlusion Ischemic Stroke

 XiaoQing Cheng,  Bing Tian, Lijun Huang, Xi Shen,  AnYu Liao,  ChangSheng Zhou,  Quan Hui Liu, HuiMin Pang, JinJing Tang, BaiYan Luo,  Xia Tian, YuXi Hou,  LuGuang Chen,  Qian Chen,  WuSheng Zhu, ChengWei Shao,  XinDao Yin, and  GuangMing Lu



## ABSTRACT

**BACKGROUND AND PURPOSE:** Early identification of malignant cerebral edema (MCE) in patients with acute ischemic stroke is crucial for timely interventions. We aimed to identify regions critically associated with MCE using the ASPECTS to evaluate the association between location-specific net water uptake (NWU) and MCE.

**MATERIALS AND METHODS:** This multicenter, retrospective cohort study included patients with acute ischemic stroke following large anterior circulation occlusion. The ASPECTS was determined by RAPID ASPECTS software. ASPECTS-NWU and Region-NWU were calculated automatically by comparing the Hounsfield unit values in the ischemic and contralateral regions. Critical ASPECTS MCE regions and Region-NWU were evaluated by multivariate logistic regression and the areas under the receiver operating characteristic curves (AUCs).

**RESULTS:** The study included 513 patients. Multivariate analysis showed that the ASPECTS insula (OR = 2.49; 95% CI, 1.44–4.31) and M5 (OR = 1.59; 95% CI, 1.11–3.41) regions were significantly associated with MCE. After adjustment, only the insula (OR = 2.34; 95% CI, 1.23–4.45) was independently associated with MCE. Univariable receiver operating characteristic curve analysis found AUCs for Insula-NWU (AUC, 0.70; 95% CI, 0.65–0.76) and ASPECTS-NWU (AUC, 0.64; 95% CI, 0.58–0.70). The Insula-NWU had better diagnostic power than ASPECTS-NWU (DeLong test;  $P = .01$ ). A multivariate regression model that combined the NIHSS, ASPECTS, insula involvement, and Insula-NWU had good discriminatory power (AUC = 0.80; 95% CI, 0.74–0.86) and better diagnostic power than Insula-NWU (DeLong test;  $P < .01$ ).

**CONCLUSIONS:** The insula region is critical for MCE, and Insula-NWU has better prediction efficacy than ASPECTS-NWU. This method does not rely on advanced imaging, facilitating rapid assessment in emergencies.

**ABBREVIATIONS:** AUC = area under the receiver operating characteristic curve; HU = Hounsfield unit; MCE = malignant cerebral edema; NWU = net water uptake; ROC = receiver operating characteristic curve

**M**alignant cerebral edema (MCE), which often occurs after an occlusion of large vessels in the anterior circulation, is a life-threatening complication. It has a prevalence of 7.5%–25.9%,<sup>1–3</sup>

Received August 14, 2024; accepted after revision December 11.

From the Department of Medical Imaging (X.C., C.Z., Q.H.L., H.P., J.T., B.L., G.L.), Department of Neurology (A.L., W.Z.), Nanjing Jinling Hospital, Nanjing University School of Medicine, Nanjing, Jiangsu, China; Department of Radiology (B.T., X.T., Y.H., L.C., C.S.), Changhai Hospital of Shanghai, The First Affiliated Hospital of Navy Medical University, Shanghai, China; Department of Medical Imaging (L.H.), the First School of Clinical Medicine, Nanjing Jinling Hospital, Southern Medical University, Nanjing, Jiangsu, China; Department of Medical Imaging (X.S.), Nanjing Jinling Hospital, Nanjing Medical University, Nanjing, Jiangsu, China; Departments of Radiology (Q.C., X.Y.), Nanjing First Hospital, Nanjing Medical University, Nanjing, Jiangsu, China.

XiaoQing Cheng and Bing Tian contributed equally to this work.

This research was funded by the National Natural Scientific Foundation of China (82271983).

Please address correspondence to XinDao Yin, MD, Department of Medical Imaging, Nanjing First Hospital, Nanjing Medical University, Nanjing, Jiangsu, China; e-mail: y.163yy@163.com



Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8659>

with death due to tissue displacement and brain herniation in nearly 80% of patients.<sup>4</sup> Early prediction of MCE in patients who could benefit from decompressive craniotomy might reduce mortality.<sup>4</sup> Results from animal studies have shown a negative correlation between x-ray attenuation and hemispheric water content. Each 1% increase in tissue water content was associated with a decrease of 1.8 Hounsfield units.<sup>5</sup> Consequently, the degree of low attenuation, as quantified by NCCT, is indicative of brain tissue water content, expressed as net water uptake (NWU). This is an increasingly used parameter to predict MCE in stroke research.<sup>6</sup>

Standardized NWU measurements require the definition of the infarct core using CTP, co-alignment of the infarct core region with the corresponding native CT, and measurement of the CT attenuation at the infarct core region and the contralateral brain tissue.<sup>6</sup> The main drawback of CTP-based NWU measurements is the time-consuming steps of image segmentation and matching, the use of which is limited to patients who have undergone CTP scans.

## SUMMARY

**PREVIOUS LITERATURE:** NWU, which quantifies the low attenuation degree on NCCT, could be replaced by the automatically measured ASPECTS-NWU to predict the development of malignant cerebral edema in ischemic stroke after acute large-vessel occlusion independent of CTP.

**KEY FINDINGS:** The insula region in ASPECTS is a strategically located region for predicting MCE. The Insula-NWU predicted MCE better than ASPECTS-NWU. Combining the NIHSS, ASPECTS, insula involvement, and Insula-NWU had a good predictive effect with an AUC of 0.80.

**KNOWLEDGE ADVANCEMENT:** The Insula-NWU broadens the usability of the automatic ASPECTS software, helps simplify the NWU measurement, and could be an important indicator in future MCE prediction models.

The ASPECTS is a semiquantitative score used to identify the extent and distribution of early ischemic changes.<sup>7</sup> The ASPECTS is divided into 10 regions based on the MCA blood supply with topologic location information. Studies have confirmed that the ASPECTS regions of infarcts at specific locations contribute differently to functional prognosis,<sup>8</sup> but the relationship with the topologic location of MCE has been rarely investigated. For each percentage increase in water content, the absolute CT attenuation varies in the white and gray matter of the brain, reflecting that different regions of brain tissue might contribute differently to the development of MCE.<sup>9</sup> It follows that the ASPECTS and NWU reflect 2 CT scan aspects of lesion progression following stroke. Whereas the ASPECTS is used to estimate the extent of the lesion, NWU can be interpreted as an indicator of the “depth of ischemia” per unit volume. The assessment of both is influenced by the location.<sup>10</sup> Automatic ASPECTS software that measures the mean CT value of the defective (ischemic) region could calculate the ASPECTS-NWU, which could then be used to predict the lesion age<sup>11</sup> and MCE.<sup>12</sup> While the 10 ASPECTS regions range in importance and size from the caudate nucleus to the M1–M6 regions, the ASPECTS-NWU calculations do not consider the relationship between the location and MCE.

We hypothesized that the ASPECTS regions of infarcts at specific locations contribute differently to MCE and that quantified location-specific NWU could contribute to the predictive power of MCE. This study aimed to do the following: 1) identify ASPECTS regions independently associated with MCE, 2) identify the relationship between location-specific NWU and MCE, and 3) determine whether location-specific NWU could differentiate MCE better than ASPECTS-NWU.

## MATERIALS AND METHODS

This study was approved by the local ethics committees of all participating institutions. The requirement for informed consent was waived by these committees. This study followed the methodology recommended in the Standards for Reporting Diagnostic Accuracy Studies (STARD) guidelines.

### Patients

This retrospective, observational cohort study assessed consecutive patients who presented with acute ischemic stroke due to occlusion of a large vessel in the anterior circulation and received reperfusion therapy between June 2021 and June 2023 at 3

centers. Inclusion criteria were as follows: 1) 18 years of age or older; 2) NCCT images obtained within 12 hours of stroke onset; 3) occlusion of a large vessel in the anterior circulation (ICA or the M1 or M2 segment of the MCA) confirmed by CTA or MRA; and 4) treatment with reperfusion therapy (endovascular thrombectomy and IV thrombolysis). Exclusion criteria included baseline NCCT showing intracranial hemorrhage or intracranial tumor, additional infarction in the posterior circulation or anterior cerebral artery region, severe artifacts or incomplete images, failure of processing by the ASPECTS software, and lack of follow-up images to assess the MCE.

We retrieved baseline variables from the electronic medical records, including demographic characteristics, vascular risk factors, medical history, laboratory and imaging data, the NIHSS on admission, and time from stroke onset to CT.

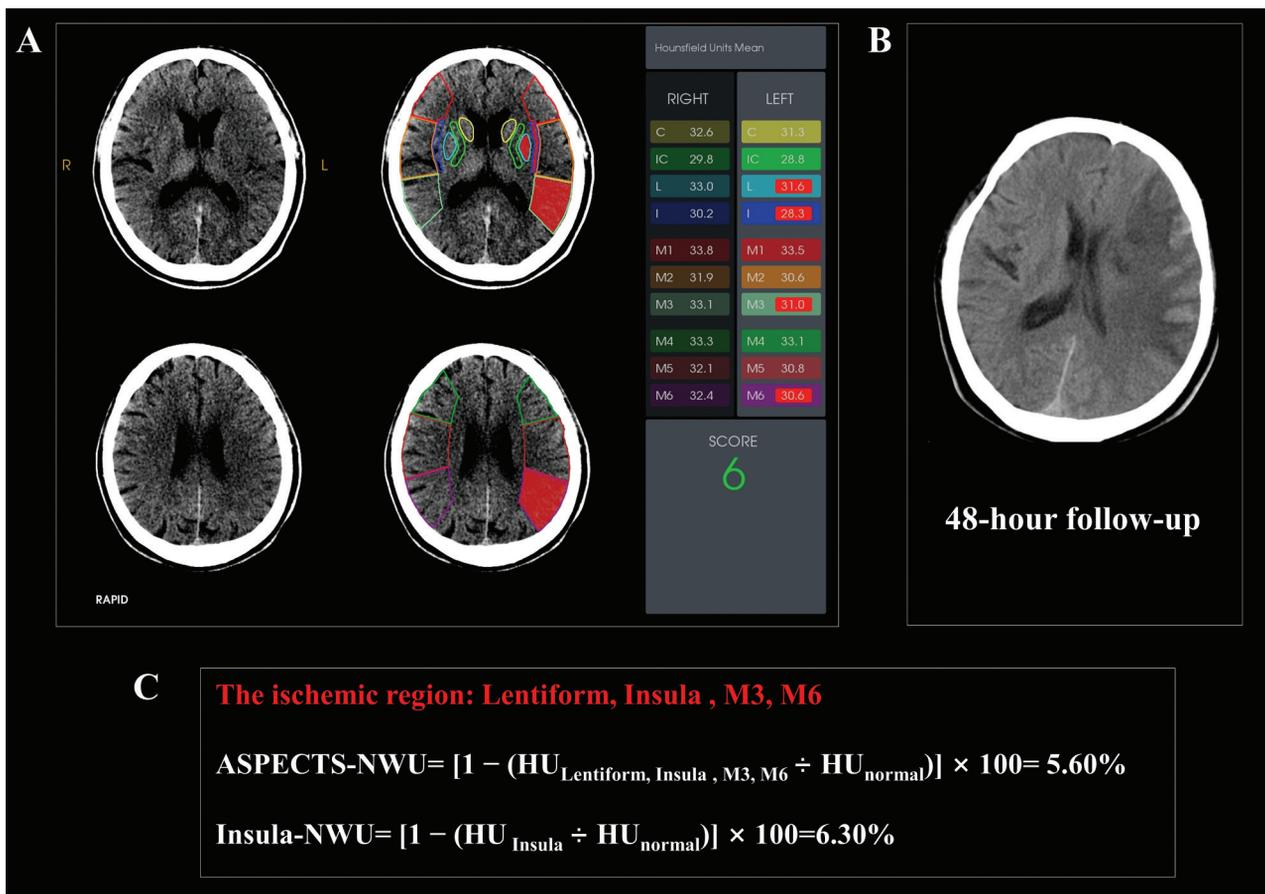
The primary outcome was MCE, defined by the presence of a large occupying infarct of the MCA causing compression of the ventricles or a midline shift of  $\geq 5$  mm on the CT on days 3–5 of follow-up, consciousness level of  $\geq 1$  on item 1a on the NIHSS, and no other causes of neurologic deterioration. The midline was determined by drawing a straight line between the anterior and posterior portions of the cerebral falx. The midline offset was quantified by drawing a second parallel line at the position where the deviation from the midline was the greatest.<sup>13</sup>

### Image Acquisition

All NCCT images from the 3 centers were in a DICOM format and generated by 1 of 4 multivector CT scanners (United Imaging uCT510, UIH; Somatom Definition, Siemens Healthineers; Sensation 64, Siemens Healthineers; Brilliance iCT, Philips Healthcare). The included NCCT images were acquired with a power of 120–40 kV, 170–380 mA, 2-second scan time, and 5-mm slice thickness.

### Image Analysis

The ASPECTS was calculated using an automatic software tool (RAPID ASPECTS, Version 4.9; iSchemaView). The 10 regions evaluated for the ASPECTS were classified into subcortical (caudate [C], lentiform [L], internal capsule [IC], and insula [I]) and superficial cortical (M1–M6) regions. First, thick-slice NCCT images (5 mm) in DICOM format were imported for preprocessing, which included removal of the skull base and cranial and CSF spaces and tilt correction of the images. A standardized atlas was then applied to create individualized grids corresponding to the 10 ASPECTS



**FIG 1.** Measurements of ASPECTS-NWU and Region-NWU by RAPID ASPECTS software. *A*, The RAPID software shows an ASPECTS score of 6. The ischemic regions were the lentiform, insula, M3, and M6. Each region had a corresponding CT value. *B*, A 48-hour follow-up CT scan shows MCE in this patient. *C*, Calculations of the ASPECTS-NWU and Region-NWU according to the Equation.  $HU_{\text{ischemic}}$  was 30.38 (the mean HU of the 4 ischemic ASPECTS regions [lentiform, insula, M3, and M6]),  $HU_{\text{normal}}$  was 32.18 (the mean HU of the respective normal ipsilateral ASPECTS regions).  $\text{Region-HU}_{\text{insula}}$  was 28.3.  $\text{Region-HU}_{\text{normal}}$  was 30.2.

regions on each hemisphere, and Hounsfield unit values were calculated for the 20 regions in the right and left brain hemispheres.

Each region was then classified as having normal or abnormal findings using a machine learning-based algorithm, and the most likely affected hemisphere was determined. Finally, an ASPECTS output map was generated, marking the affected region in red and reporting the ASPECTS.

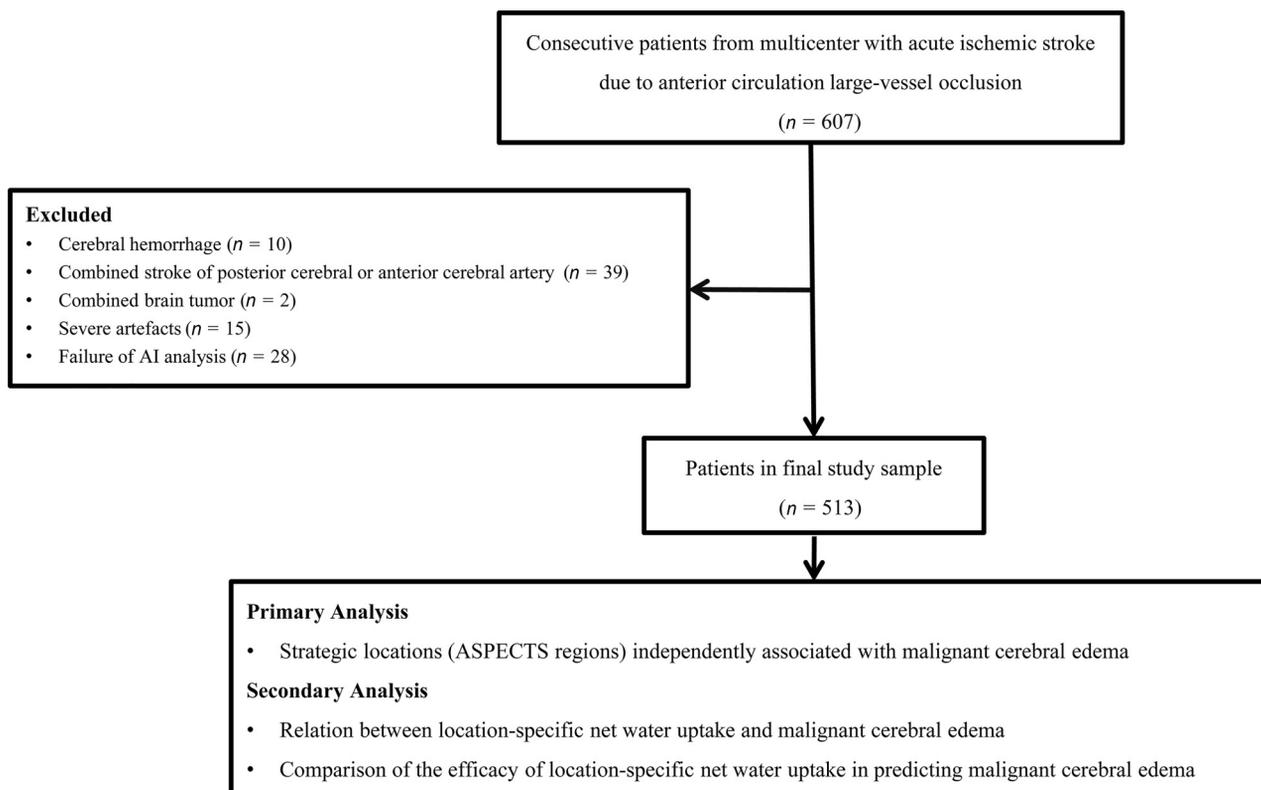
On the basis of the regional Hounsfield unit (HU) obtained by the RAPID ASPECTS software, we calculated the mean HU values for the ischemic ( $HU_{\text{ischemic}}$ ) and corresponding contralateral ( $HU_{\text{normal}}$ ) regions and determined the NWU value for each according to the Equation. The ASPECTS-NWU was calculated using the average HU values for the overall ischemic and corresponding contralateral regions based on the ischemic region obtained from the RAPID ASPECTS software (Fig 1).<sup>11</sup> Region-NWU was calculated using the HU values for individual ischemic and contralateral ASPECTS regions (Fig 1).

$$\text{NWU}\% = [1 - (HU_{\text{ischemic}} \div HU_{\text{normal}})] \times 100.$$

### Statistical Analysis

Data were analyzed using statistical packages in R (R Foundation; <http://www.r-project.org>; Version 3.4.3) and Empower-Stats

(<http://www.empowerstats.com>, X&Y Solutions). Continuous variables with a normal distribution are presented as means, while categorical variables are reported as No. (%). Baseline characteristics and stroke information in the MCE and no-MCE groups were compared using the  $\chi^2$  test for categorical variables and the 1-way ANOVA and Kruskal-Wallis tests for continuous variables with normal and skewed distributions, respectively. Multivariate logistic regression analysis was used to assess the association between the ASPECTS regions and MCE. The NWU was calculated separately for individual ASPECTS regions and the entire ASPECTS ischemic region (ASPECTS-NWU). We report ORs with 95% CIs. Smoothed spline plots were created to graphically depict the link between the variables and the MCE. Receiver operating characteristic curve (ROC) analysis was used to assess the diagnostic performance and sensitivity and specificity metrics of variables such as Region-NWU, ASPECTS-NWU, and baseline NIHSS scores in predicting MCE. We generated the 95% CIs for the areas under the receiver operating characteristic curve (AUCs) using the exact method, also known as the DeLong method. The optimal cut-point is determined by the threshold that maximizes the Youden index, effectively balancing sensitivity and specificity. Statistical methods for model comparison were performed by the DeLong method, with *P* values corrected for Bonferroni correction and statistically significant at .025.



**FIG 2.** Flow chart of patient selection for the study.

## RESULTS

### Patient Characteristics

The study flow chart is presented in Fig 2. Among the 607 consecutively included patients with acute ischemic stroke due to acute large-vessel occlusion in the anterior circulation, 94 were considered ineligible for the study: 10 with cerebral hemorrhage on baseline NCCT scans; 2 with a brain tumor; 39 with a vertebral basilar, posterior cerebral, or anterior cerebral artery regional stroke; 15 with severe artifacts; and 28 with software analysis failure. Finally, 513 patients (314 men; mean age, 67.4 [SD, 12.3] years) were included in the study. The baseline clinical characteristics of the patients are shown in the Table.

MCE was confirmed in follow-up CT in 116 (22.6%) patients. A comparison of data between patients with and without MCE is shown in the Table. Patients with MCE had higher NIHSS scores at admission (mean, 18.12 versus 13.94;  $P < .01$ ), ASPECTS-NWU values (mean, 7.20 versus 5.49;  $P < .01$ ), and blood glucose (mean, 8.20 versus 7.44 mmol/L;  $P = .01$ ), and lower ASPECTS (mean, 4.52 versus 6.58;  $P < .01$ ). Differences in all ASPECTS regions were statistically significant.

### Primary Analysis: Analyses of the Association between the Region and MCE

Multivariate logistic regression analysis showed that the regions significantly associated with MCE were M5 (OR, 1.95; 95% CI, 1.11–3.41;  $P = .02$ ) and the insula (OR, 2.49; 95% CI, 1.44–4.31;  $P = .01$ ). The association with the remaining 8 regions (M1, M2, M3, M4, M6, lentiform nucleus, internal capsule, and

caudate) did not reach statistical significance (Supplemental Data). After we adjusted for sex, age, NIHSS, and ASPECTS, the analysis showed that the insula region remained independently correlated with MCE (OR, 2.34; 95% CI, 1.23–4.45;  $P = .01$ ), but the association with the M5 region was statistically insignificant (OR, 1.54; 95% CI, 0.80–2.97; Supplemental Data).

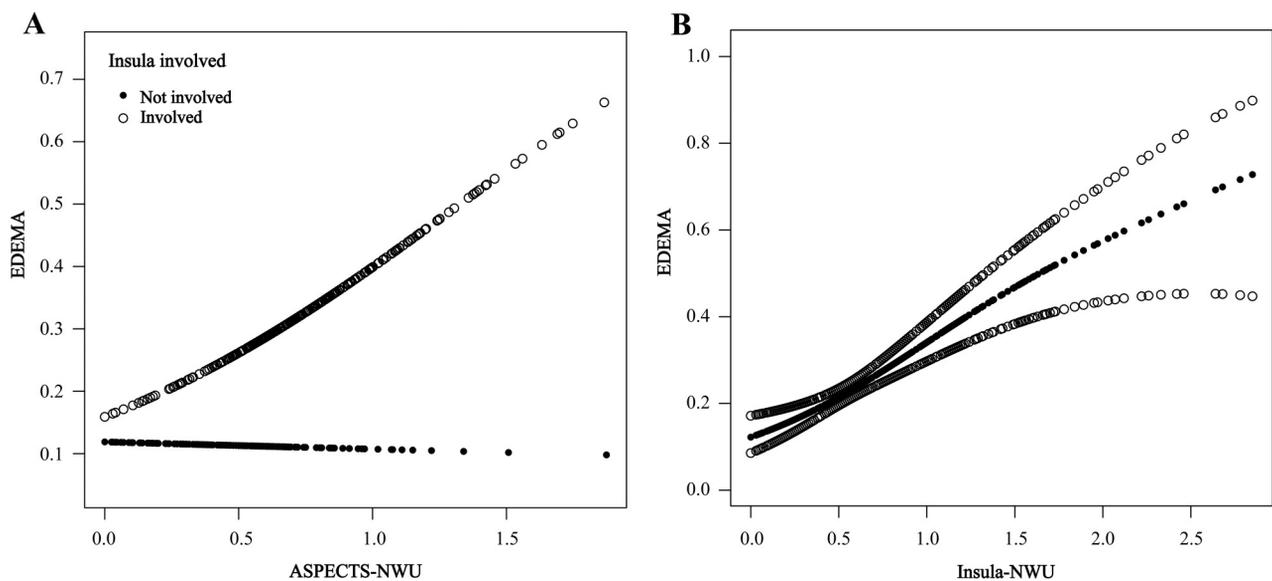
### Secondary Analysis: Association between the Region-NWU and MCE

Figure 3A shows the relationship between ASPECTS-NWU and the MCE risk, stratified by insula involvement. When infarcts involved the insula, the ASPECTS-NWU increased and the risk of MCE was elevated, whereas if infarcts did not involve the insula, this association was absent. Figure 3B shows the association of insula-NWU with MCE.

Because multivariate logistic regression analysis found that the insula and M5 regions were independently associated with MCE, they were combined to calculate the Insula+M5-NWU. Univariable ROC analysis found AUCs for Insula-NWU (AUC, 0.70; 95% CI, 0.65–0.76), Insula + M5-NWU (AUC, 0.71; 95% CI, 0.65–0.76), ASPECTS-NWU (AUC, 0.64; 95% CI, 0.58–0.70), and NIHSS (AUC, 0.70; 95% CI, 0.65–0.75). The diagnostic power of Insula-NWU was statistically better than that of ASPECTS-NWU (DeLong test;  $P = .01$ ) (Fig 4). The optimal cutoff for Insula-NWU to classify MCE, the NWU value with the highest Youden index, was 5.75 (sensitivity 68.10%; specificity 66.50%). A model developed with NIHSS, ASPECTS,

### Clinical and imaging characteristics of the patient population

Variable	All	MCE-	MCE+	P Value
Age (mean) (yr)	67.43 (SD 12.33)	67.10 (SD 12.55)	68.54 (SD 11.55)	.27
Male sex, No. (%)	314 (61%)	241 (61%)	73 (63%)	.67
NIHSS score at baseline (mean)	14.88 (SD 6.57)	13.94 (SD 6.53)	18.12 (SD 5.63)	<.01
Time from stroke onset to CT (mean) (h)	4.60 (SD 4.23)	4.63 (SD 4.47)	4.49 (SD 3.31)	.77
ASPECTS (mean)	6.11 (SD 2.83)	6.58 (SD 2.57)	4.52 (SD 3.09)	<.01
ASPECTS-NWU (mean)	5.87 (SD 3.87)	5.49 (SD 3.79)	7.20 (SD 3.85)	<.1
Blood glucose (mmol/L) (mean)	7.6 (SD 2.8)	7.44 (SD 2.77)	8.20 (SD 3.04)	.01
Risk factors				
Diabetes mellitus, No. (%)	125 (24%)	100 (25%)	25 (22%)	.41
Hypertension, No. (%)	329 (64%)	252 (64%)	77 (66%)	.59
Previous ischemic stroke or TIA, No. (%)	101 (20%)	76 (19%)	25 (22%)	.57
Atrial fibrillation, No. (%)	167 (33%)	126 (32%)	41 (35%)	.47
Coronary artery disease, No. (%)	67 (13%)	50 (13%)	17 (15%)	.56
Occlusion site				
ICA, No. (%)	158 (38%)	109 (35%)	49 (50%)	.01
MCA-M1 segment, No. (%)	231 (56%)	184 (58%)	47 (48%)	
MCA-M2 segment, No. (%)	24 (6%)	22 (7%)	2 (2%)	
ASPECTS region, No. (%)				
M1	191 (37%)	126 (32%)	65 (56%)	<.01
M2	182 (35%)	119 (30%)	63 (54%)	<.01
M3	141 (27%)	90 (23%)	51 (44%)	<.01
M4	140 (27%)	94 (24%)	46 (40%)	<.01
M5	172 (34%)	108 (27%)	64 (55%)	<.01
M6	135 (26%)	87 (22%)	48 (41%)	<.01
Insula	285 (56%)	195 (49%)	90 (78%)	<.01
Lentiform nucleus	248 (48%)	176 (44%)	72 (62%)	<.01
Internal capsule	231 (45%)	164 (41%)	67 (58%)	.02
Caudate nucleus	264 (51%)	194 (49%)	70 (60%)	.03

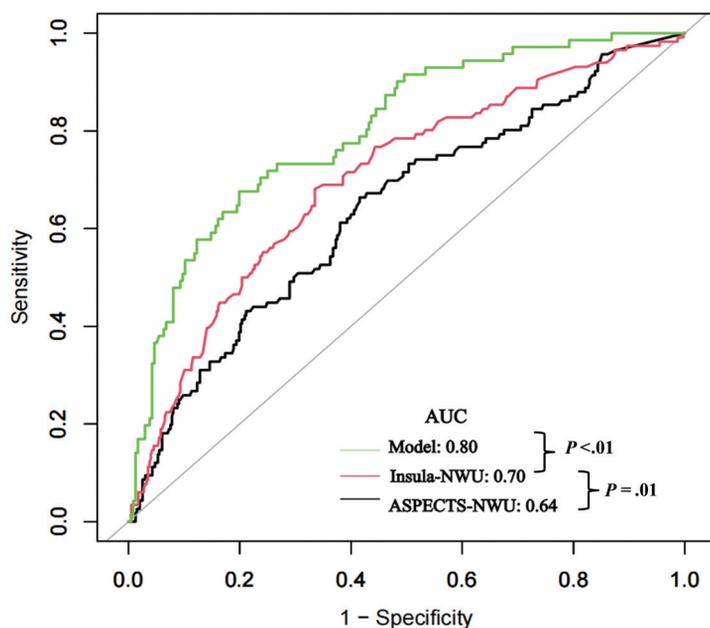


**FIG 3.** The smoothing spline plots show the relationship between NWU and MCE. A, The relationship between ASPECTS-NWU and MCE risk, stratified by insula involvement. When infarcts involved the insula (*hollow circles*), ASPECTS-NWU and the risk of MCE increased, whereas when the infarcts did not involve the insula (*black circles*), this association was absent. B, The association between the Insula-NWU and MCE (*black circle dotted line*). The area between the 2 *hollow circle lines* is the 95% CI.

insula involvement, and Insula-NWU showed good discriminatory power with an AUC of 0.80 (95% CI, 0.74–0.86), sensitivity of 67.61%, and specificity of 80.08%. The diagnostic power of the model was statistically better than that of Insula-NWU (DeLong test;  $P < .01$ ) (Fig 4).

### DISCUSSION

This study aimed to identify the strategic locations associated with MCE and the efficacy of Region-NWU in predicting MCE in patients with acute ischemic stroke due to large-vessel occlusion in the anterior circulation. Our main finding was that the



	AUC	95% CI	Cut-off	Sensitivity	Specificity
ASPECTS-NWU	0.64	0.58-0.70	6.07	66.38	58.44
Insula-NWU	0.70	0.65-0.76	5.75	68.10	66.50
<b>Model</b> (NIHSS, ASPECTS, Insula involvement, and Insula-NWU)	0.80	0.74-0.86	-	67.61	80.08

**FIG 4.** ROC analysis resulted in AUCs of 0.64 and 0.70 for ASPECTS-NWU and Insula-NWU, respectively. The multivariate regression model combining NIHSS, ASPECTS, insula involvement, and Insula-NWU resulted in an AUC of 0.80. The diagnostic power of Insula-NWU is superior to that of ASPECTS-NWU by the DeLong test ( $P = .01$ ), and the diagnostic power of the model is superior to that of Insula-NWU ( $P < .01$ ).

insula and M5 ASPECTS regions, especially the insula, were independently associated with MCE. Furthermore, Region-NWU was a more valid and simple predictor of MCE than the ASPECTS-NWU. Simplified NWU measurements could be made automatically with ASPECTS software, eliminating the need for advanced imaging and complex postprocessing and making it attractive for future applications.

The present study found that the insula was strongly associated with the development of MCE. The ASPECTS regions of infarcts at specific locations have been shown to contribute differently to functional prognosis. A recent meta-analysis found that infarcts were unevenly distributed across the ASPECTS regions and might be unevenly weighted in predicting the prognosis of patients with acute ischemic stroke, with infarcts in the M6 region being the strongest predictor.<sup>8</sup> A study of the relationship between specific ASPECTS regions and good functional outcomes showed that the insula (OR, 0.56; 95% CI, 0.42–0.75) and M5 (OR, 0.53; 95% CI, 0.29–0.97) regions were negatively correlated with good functional outcomes and had the strongest effect.<sup>14</sup>

However, there are only a few studies on important positional features associated with MCE. A study based on voxel-based lesion-symptom mapping found that the lesion distribution in

the temporal and frontal lobes was mildly-to-moderately predictive of the need for a decompressive craniectomy.<sup>15</sup> However, the small sample of that study and the cumbersome postprocessing required when using the voxel-based lesion-symptom mapping approach make it difficult to apply to individual patients. Another study used the ASPECTS on follow-up CT, combining M2, M3, and the insula as the middle fossa. For patients without mechanical thrombectomy, the middle cranial fossa (OR, 2.57; 95% CI, 1.12–6.25) was independently associated with potentially lethal malignant edema.<sup>3</sup> Our study showed that the insular region had good clinical interpretability of MCE. The insula is adjacent to the frontal/temporal opercula at the brain surface and medial to the Sylvian fissure and is directly supplied by the proximal portion of the 2 main MCA M2 branches after they branch at a right angle from the mainstem (M1), a topographic feature that predisposes them to embolisms, especially cardiac emboli, resulting in occlusion of the region.<sup>16</sup>

It has been reported that insula involvement might indicate a higher risk of conversion of salvageable penumbra to irreversibly damaged tissue,<sup>17</sup> which is associated with a poorer

clinical prognosis. In addition, co-involvement of the insula and M5 regions might have a more significant mass effect on the midline structure. Moreover, insula dysfunction might be directly or indirectly associated with stroke complications such as increased permeability of the BBB, hospital-acquired pneumonia, cardiac arrhythmias, and hyperglycaemia.<sup>18,19</sup> The present study provides further evidence of the importance of insula infarction for the development of MCE.

Another important point was that the present study further combined the critical regions with NWU to reflect the relationship between the water content of the brain tissue in these regions and the occurrence of MCE. Broocks et al<sup>20</sup> showed that it was possible to quantify NWU in infarcted lesions by measuring relative CT attenuation. They proposed NWU as an important alternative imaging biomarker for MCE in anterior circulation stroke. A study in patients with posterior circulation stroke similarly demonstrated a high discriminatory power for the occurrence of MCE when the NWU was >14.9% (AUC of 0.94).<sup>21</sup> However, the standard procedure for measuring NWU in the above studies included CTP to ensure a precise definition of the core lesions used for attenuation measurements, limiting the clinical application of NWU. The ASPECTS-NWU, automatically measured by ASPECTS scoring software, is independent of CTP

and could be used as an alternative measurement approach.<sup>11,12</sup> However, it does not consider differences in importance among the ASPECTS regions and the varying degrees of water uptake among brain tissues.

The strength of this study is that it considers a strategically important region and combines NWU with it to yield the Insula-NWU with significantly better efficacy than ASPECTS-NWU in predicting MCE, further simplifying the assessment process. The combination of admission NIHSS, ASPECTS, insula involvement, and Insula-NWU resulted in a predictive AUC of >0.8. Therefore, Insula-NWU could be an important metric for subsequent predictive models. Moreover, it broadens the usability of the automatic ASPECTS software and helps simplify NWU measurements.

Our study had some limitations. First, it focused on the brain region significantly associated with MCE and the NWU in this region; however, the clinical prognosis of the patients was not analyzed. Related work will be performed in subsequent studies. Second, the NWU obtained on the basis of automatic ASPECTS software or CTP differs in measurement and cutoff values when predicting MCE, and this study could not compare the NWU values measured by the 2 methods. Third, the model in this study obtained the prediction results of MCE through multivariate logistic regression analysis, and it is possible that the results of the model were overly optimistic due to the lack of cross-validation.

## CONCLUSIONS

The occurrence of MCE after acute ischemic stroke due to acute large-vessel occlusion is location-specific. The insula, a critical location, was combined with NWU to obtain a quantitative variable that could predict the occurrence of MCE. This approach does not rely on advanced imaging modalities, simplifying the MCE assessment and facilitating rapid assessment in emergency situations.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Zhang X, Yan S, Zhong W, et al. **Early NT-ProBNP (N-Terminal Probrain Natriuretic Peptide) elevation predicts malignant edema and death after reperfusion therapy in acute ischemic stroke patients.** *Stroke* 2021;52:537–42 CrossRef Medline
2. Huang X, Chen C, Wang H, et al. **The ACORNS grading scale: a novel tool for the prediction of malignant brain edema after endovascular thrombectomy.** *J Neurointerv Surg* 2023;15:e190–97 CrossRef Medline
3. Stafford R, Chatzidakis S, Kim IS, et al. **Follow-up ASPECTS improves prediction of potentially lethal malignant edema in patients with large middle cerebral artery stroke.** *J Neurointerv Surg* 2024;17:e83–86 CrossRef Medline
4. Kimberly WT, Dutra BG, Boers AM, et al; MR CLEAN Investigators. **Association of reperfusion with brain edema in patients with acute ischemic stroke: a secondary analysis of the MR CLEAN Trial.** *JAMA Neurol* 2018;75:453–61 CrossRef Medline

5. Dzialowski I, Weber J, Doerfler A, et al. **Brain tissue water uptake after middle cerebral artery occlusion assessed with CT.** *J Neuroimaging* 2004;14:42–48 Medline
6. Minnerup J, Broocks G, Kalkoffen J, et al. **Computed tomography-based quantification of lesion water uptake identifies patients within 4.5 hours of stroke onset: a multicenter observational study.** *Ann Neurol* 2016;80:924–34 CrossRef Medline
7. Barber PA, Demchuk AM, Zhang J, et al. **Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy: ASPECTS Study Group: Alberta Stroke Programme Early CT Score.** *Lancet* 2000;355:1670–74 CrossRef Medline
8. Seyedasadat SM, Neuhaus AA, Pederson JM, et al. **Location-specific ASPECTS paradigm in acute ischemic stroke: a systematic review and meta-analysis.** *AJNR Am J Neuroradiol* 2020;41:2020–26 CrossRef Medline
9. Vorasayan P, Bevers MB, Beslow LA, et al. **Intravenous glibenclamide reduces lesional water uptake in large hemispheric infarction.** *Stroke* 2019;50:3021–27 CrossRef Medline
10. Broocks G, Meyer L, Hanning U, et al. **Haemorrhage after thrombectomy with adjuvant thrombolysis in unknown onset stroke depends on high early lesion water uptake.** *Stroke Vasc Neurol* 2024;9:390–98 CrossRef Medline
11. Cheng X, Wu H, Shi J, et al. **ASPECTS-based net water uptake as an imaging biomarker for lesion age in acute ischemic stroke.** *J Neurol* 2021;268:4744–51 CrossRef Medline
12. Shi J, Wu H, Dong Z, et al. **Automated quantitative lesion water uptake in acute stroke is a predictor of malignant cerebral edema.** *Eur Radiol* 2022;32:2771–80 CrossRef Medline
13. Irvine HJ, Ostwaldt AC, Bevers MB, et al. **Reperfusion after ischemic stroke is associated with reduced brain edema.** *J Cereb Blood Flow Metab* 2018;38:1807–17 CrossRef Medline
14. Neuberger U, Vollherbst DF, Ulfert C, et al. **Location-specific ASPECTS does not improve outcome prediction in large vessel occlusion compared to cumulative ASPECTS.** *Clin Neuroradiol* 2023;33:661–68 CrossRef Medline
15. McCullough-Hicks M, Topiwala K, Christensen S, et al. **Anatomical predictors of need for decompressive craniectomy after stroke using voxel-based lesion symptom mapping.** *J Neuroimaging* 2023;33:737–41 CrossRef Medline
16. Regenhardt RW, Singhal AB, He J, et al. **Percent insular ribbon infarction for predicting infarct growth rate and 90-day outcomes in large-vessel occlusive stroke: secondary analysis of prospective clinical trial data.** *AJR Am J Roentgenol* 2023;221:103–13 CrossRef Medline
17. Kamalian S, Kemmling A, Borgie RC, et al. **Admission insular infarction >25% is the strongest predictor of large mismatch loss in proximal middle cerebral artery stroke.** *Stroke* 2013;44:3084–89 CrossRef Medline
18. Raichle ME, Hartman BK, Eichling JO, et al. **Central noradrenergic regulation of cerebral blood flow and vascular permeability.** *Proc Natl Acad Sci U S A* 1975;72:3726–30 CrossRef Medline
19. Tokunaga K, Toyoda K, Kimura S, et al; SAMURAI-NVAF and RELAXED Investigators. **Association of the timing of atrial fibrillation detection and insular involvement with the risk of embolic events after acute ischemic stroke.** *Ann Neurol* 2024;95:338–46 CrossRef Medline
20. Broocks G, Flottmann F, Ernst M, et al. **Computed tomography-based imaging of voxel-wise lesion water uptake in ischemic brain: relationship between density and direct volumetry.** *Invest Radiol* 2018;53:207–13 CrossRef Medline
21. Broocks G, Elsayed S, Kniep H, et al. **Early prediction of malignant cerebellar edema in posterior circulation stroke using quantitative lesion water uptake.** *Neurosurgery* 2021;88:531–37 CrossRef Medline

# A Method for Imaging the Ischemic Penumbra with MRI Using Intravoxel Incoherent Motion

Mira M. Liu, Niloufar Saadat, Steven P. Roth, Marek A. Niekrasz, Mihai Giurcanu, Mohammed Salman Shazeeb, Timothy J. Carroll, and Gregory A. Christoforidis



## ABSTRACT

**BACKGROUND AND PURPOSE:** In acute ischemic stroke, the amount of “local” CBF distal to the occlusion, ie, all blood flow, whether supplied antegrade or delayed and dispersed through the collateral network, may contain valuable information regarding infarct growth rate and treatment response. DSC processed with a local arterial input function (AIF) is one method of measuring local CBF (local-qCBF) and has been shown to correlate with collateral supply. Similarly, intravoxel incoherent motion MRI (IVIM) is “local,” with excitation and readout in the same plane, and a potential alternative way to measure local-qCBF. This work compares IVIM local-qCBF against DSC local-qCBF in the ischemic penumbra, compares the measurement of perfusion-diffusion mismatch (PWI/DWI), and examines if local-qCBF may improve prediction of the final infarct.

**MATERIALS AND METHODS:** Eight experiments in a preclinical canine model of middle cerebral artery occlusion were performed. Native collateral circulation was quantified via x-ray DSA 30 minutes postocclusion, and collateral supply was subsequently enhanced in a subset of experiments with simultaneous pressor and vasodilator. IVIM, DSC, and DWI MRI were acquired 2.5 hours postocclusion. IVIM was postprocessed to return local-qCBF from  $fD^*$ , water transport time (WTT) from  $D^*$ , diffusion from  $D$ , and the PWI/DWI mismatch. These were compared with DSC parameters processed first with a standard global-AIF and then with a local-AIF. These DSC parameters included time-to-maximum, local MTT, standard-qCBF, local-qCBF, and PWI/DWI mismatch. Infarct volume was measured with DWI at 2.5 hours and 4 hours postocclusion.

**RESULTS:** Two and one-half hours postocclusion, IVIM local-qCBF in the noninfarcted ipsilateral territory correlated strongly with DSC local-qCBF (slope = 1.00,  $R^2 = 0.69$ , Lin CCC = 0.77). Correlation was weaker between IVIM local-qCBF and DSC standard-qCBF ( $p = 0.38$ ,  $R^2 = 0.13$ ). DSC local-qCBF and IVIM local-qCBF in the noninfarcted ipsilateral territory both returned strong prediction of final infarct volume ( $R^2 = 0.78$ ,  $R^2 = 0.61$ , respectively). DSC standard-qCBF was a weaker predictor ( $R^2 = 0.12$ ). The hypoperfused lesion from DSC local-qCBF and from IVIM local-qCBF both predicted final infarct volume with good sensitivity and correlation (slope = 2.08,  $R^2 = 0.67$ , slope = 2.50,  $R^2 = 0.68$ , respectively). The IVIM PWI/DWI ratio was correlated with infarct growth ( $R^2 = 0.70$ ), and WTT correlated with DSC MTT (slope = 0.82,  $R^2 = 0.60$ ).

**CONCLUSIONS:** Noncontrast IVIM measurement of local-qCBF and PWI/DWI mismatch may include collateral circulation and improve prediction of infarct growth.

**ABBREVIATIONS:** AIF = arterial input function; Gd = gadolinium; IR = inversion recovery; IVIM = intravoxel incoherent motion; MCAO = middle cerebral artery occlusion; MD = mean diffusivity; qCBF = quantitative cerebral blood flow; Tmax = time to maximum; WTT = water transport time

Acute ischemic stroke occurs when sudden loss of blood supply to a region of the brain leads to damaged or dead brain cells. Rapid early intervention by thrombectomy or thrombolysis

after symptom onset<sup>1</sup> is most effective in treating stroke.<sup>2,3</sup> However, large multicenter trials<sup>4-6</sup> have suggested that the treatment window can be extended for some patients. This “late thrombectomy” can be performed 6 to 24 hours after symptom

Received November 11, 2024; accepted after revision January 2, 2025.

From the Departments of Radiology Medical Physics (M.M.L., T.J.C.), Interventional Radiology (N.S., G.A.C.), Anesthesiology (S.P.R.), Surgery and Large Animal Studies (M.A.N.), and Statistics (M.G.), University of Chicago, Chicago, Illinois; Biomedical Engineering and Imaging Institute (M.M.L.), Icahn School of Medicine at Mount Sinai, New York, New York; Radiology (M.S.S.), University of Massachusetts Chan Medical School, Worcester, Massachusetts; and Department of Radiology (G.A.C.), Mount Carmel Health Systems, Columbus, Ohio.

This research has been supported by the US National Institutes of Health R01-NS093908 (Carroll/Christoforidis) and The National Science Foundation DGE-1746045 (Liu).

Please address correspondence to Mira M. Liu, Icahn School of Medicine at Mount Sinai, BioMedical Engineering and Imaging Institute, 1470 Madison Ave, 1st Floor, NY, NY 10029; e-mail: liusarkarm@uchicago.edu



Indicates article with supplemental data.



This article contains an imaging protocol available on our website. <http://dx.doi.org/10.3174/ajnr.A8656>

## SUMMARY

**PREVIOUS LITERATURE:** In ischemic stroke, the “local” nature of IVIM with excitation and readout in the same plane allows it to image all capillary motion. This includes blood flow from the collateral network, even if it has arrived delayed and dispersed relative to antegrade flow. Similarly, correcting DSC quantitative CBF (qCBF) for this delay and dispersion by using a local AIF improves agreement with collateralization. While IVIM in stroke has predominately been compared with standard DSC, comparison of IVIM local-qCBF to DSC local-qCBF may improve agreement and demonstrate how local-qCBF could add valuable information to stroke imaging.

**KEY FINDINGS:** In a canine model of middle cerebral artery occlusion, IVIM local-qCBF in the ipsilateral MCA territory correlated strongly with DSC local-qCBF ( $R^2 = 0.69$ ), and weakly with DSC standard-qCBF ( $R^2 = 0.16$ ). Local-qCBF returned a stronger prediction of final infarct volume (DSC  $R^2 = 0.78$ , IVIM  $R^2 = 0.61$ ) than standard-qCBF ( $R^2 = 0.12$ ). IVIM PWI/DWI ratio was strongly correlated with infarct growth ( $R^2 = 0.70$ ).

**KNOWLEDGE ADVANCEMENT:** These findings support 1) IVIM as a noncontrast method of measuring local-qCBF, including collateral circulation, in acute stroke, 2) improved agreement of IVIM and DSC with a local-AIF, and 3) using simultaneous perfusion-weighted and diffusion-weighted images from IVIM for PWI/DWI mismatch.

onset, which is a crucial extension for patients with unknown onset time. Patient eligibility for late thrombectomy may be determined by the “mismatch” between 1) the hypoperfused lesion distal to an occlusion imaged with PWI and 2) the amount of irretrievable, dead/infarcted brain imaged with DWI.<sup>7,8</sup> This PWI-DWI mismatch is one measure of the penumbra, which may be useful for patient triage and novel stroke therapeutic studies.<sup>9-13</sup>

Hypoperfusion can be measured as prolonged time-to-maximum (Tmax) or delayed arrival time of the contrast bolus.<sup>14</sup> However, even when delayed bolus makes regions of the brain appear hypoperfused, some of these regions may be sustained by collateral circulation.<sup>15,16</sup> “Local” CBF distal to the occlusion, ie, all blood flow, whether supplied antegrade or delayed and dispersed through the collateral network, is one way of measuring this collateral circulation. Including collateral circulation in perfusion imaging may provide more complete pathophysiologic information for each patient, improve patient selection for reperfusion therapy, and better predict infarct growth.<sup>15-17</sup> Recently, DSC CBF with a local arterial input function (AIF)<sup>20</sup> has been shown to measure local quantitative CBF (local-qCBF in mL/100 g/min) that correlates with collateralization and infarct growth better than standard DSC.<sup>18</sup> Similarly, intravoxel incoherent motion (IVIM) MRI perfusion fraction has shown promise as a method of imaging collateral flow.<sup>19</sup> As such, imaging local-qCBF with IVIM may add valuable information about collateral circulation, hypoperfusion in the ischemic penumbra, and infarct growth.

This work aims to examine the use of noncontrast IVIM to measure local-qCBF and measure the ischemic penumbra in acute ischemic stroke. To study the nature of the local flow, IVIM parameters are compared with both local-AIF DSC and standard DSC parameters in a preclinical canine model of acute middle cerebral artery occlusion (MCAO). First, it compares IVIM local-qCBF to DSC local-qCBF and DSC standard-qCBF in the ischemic penumbra. Second, it compares the predictive ability of hypoperfusion lesions from local-qCBF versus standard-qCBF against final infarct. Finally, it uses IVIM simultaneous local-qCBF PWI and DWI ratio to predict infarct growth and

compares it to the PWI/DWI ratio from DSC Tmax, DSC local-qCBF, and DSC standard-qCBF.

## MATERIALS AND METHODS

### Theory

Contrast bolus is delayed and dispersed when traveling around an occlusion through the collateral network. Mathematically, standard DSC perfusion analysis may not include compensatory blood that travels through collaterals if this delay is not corrected.<sup>20</sup> This means that standard DSC without correction may overestimate the hypoperfusion lesion when there are good collaterals. Use of a voxelwise local-AIF has been shown to correct this delay.<sup>21,22</sup> With a local-AIF generated for every voxel, rather than a standard single global-AIF in a brain-feeding artery, DSC local-qCBF captures all of the perfusion in a voxel, including collateral circulation.<sup>18</sup>

In theory, IVIM local-qCBF in the penumbra should more closely agree with DSC local-qCBF than standard DSC, especially in cases with good collateral circulation. This is because IVIM is local, with excitation and readout on the same plane.<sup>24</sup> It measures capillary flow independent of contrast bolus arrival by separating intravoxel motion into blood pseudo-diffusion and tissue diffusion.<sup>23</sup> Both Local-AIF DSC and IVIM should be sensitive to all the blood in capillaries, whether supplied antegrade or retrograde through the collateral network.<sup>18,19</sup> The speed of the capillary motion can also estimate how long it would take for 50% of the original molecules in a volume to diffuse out<sup>24</sup>; this theoretically could be comparable to local-AIF DSC mean transit time. Further details of the theory behind local-qCBF are available in the Supplemental Data.

### Preclinical Canine Model

All experiments were conducted by using a previously reported preclinical canine model of ischemic stroke.<sup>25</sup> The 2-day experimental protocol was approved by the University of Chicago Institutional Animal Care and Use Committee and reported in compliance with ARRIVE guidelines. The University of Chicago is an AAALAC International accredited institution adhering to

the following guidelines, regulations, and policies: a) Guide for the Care and Use of Laboratory Animals (National Research Council), b) USDA Animal Welfare Act and Animal Welfare Regulations, and c) Public Health Service Policy on Humane Care and Use of Laboratory Animals.

Eight canines (mean age =  $3.4 \pm 3.9$  years, mean weight =  $25.3 \pm 5.0$  kg, 7 female, 1 male) underwent permanent endovascular MCAO via embolic occlusion coils under fluoroscopic guidance with R/L randomization. M1 occlusion was verified via selective internal carotid and vertebral arteriography.<sup>25</sup> Native collateral circulation was quantified 30 minutes post-MCAO by assessing x-ray arteriographic images (OEC9800; GE Healthcare) with a pial collateral score<sup>26</sup> modified for canines as described previously.<sup>16</sup> Further details of the canine model and collateral scoring are available in the Supplemental Data. To evaluate a larger range of collateral supply, 5 of 8 subjects underwent flow augmentation (simultaneous pressor and vasodilator norepinephrine and hydralazine); treated and untreated subjects were pooled. In previous work, this flow augmentation has been shown to increase collateral circulation and slow infarct growth by disruption of cerebral autoregulation of poor collaterals.<sup>13,27</sup>

### **MRI Acquisition**

MRI was acquired 2.5 hours after occlusion with a 3T MRI scanner (Ingenia, Philips) head-first, prone position with a 15-channel receive-only coil. Sequences were acquired with DSC, followed by DTI, and then IVIM.

DSC images (coronal plane, 2D gradient-echo, T2\*-weighted EPI, FOV/matrix = 160 mm/176, 5 slices/6 mm thick, TR/TE = 500/30, 120 phases, total scan time = 60 seconds) were taken along with rapid 15-second T1 maps following the "T1-bookend method"<sup>21,22,28-34</sup> (2D inversion recovery [IR] Look-Locker, single-shot EPI FOV/matrix = 160 mm/176, 5 slices/6 mm thick). Gadolinium (Gd)-based contrast agent (Multihance, Bracco) followed by a saline flush (Gd: 3 mL at 2 mL/s, saline: 20 mL at 2 mL/s) was injected in the forepaw.

DTI for mean diffusivity (MD), analogous to more widely used ADC, was acquired to measure infarct volume every 30 minutes post-MCAO and track infarct growth over time.<sup>16</sup> A stack of 50 2D DTI slices was prescribed to cover the entire head (slice thickness = 2 mm, FOV =  $128 \times 128$  mm/matrix =  $128 \times 128$ , TR/TE = 2993/83 ms, flip angle = 90°, b-values = 0, 800 seconds/mm<sup>2</sup>, 32 directions).

Multi b-value DWI for IVIM was collected with 10 b-values (0, 111, 222, 333, 444, 556, 667, 778, 889, 1000 seconds/mm<sup>2</sup>) and 3 orthogonal directions. Scans covered the entire head (2D single-shot EPI, TR/TE = 3056/91 ms, 50 slices/2 mm thick, FOV/matrix = 224 mm/128 or FOV/matrix 160 mm/96, total scan time = 5.5 minutes, SENSE Factor = 2). The MR protocol is available in the Supplemental Data.

### **DSC Analysis**

DSC local-qCBF was postprocessed in mL/100 g/min by using the well-established T1-bookend method<sup>21,22,28-34</sup> with a voxel-by-voxel local-AIF. This local-AIF corrected for the late arrival (delay) and "blunting" (dispersion) of the contrast bolus as it propagated through the brain<sup>22</sup> and included collateral supply in the ischemic penumbra.<sup>18</sup>

DSC was also processed with a standard single global-AIF and quantified as standard-qCBF with the same T1-bookend method. Tmax was calculated as the time at which the residue function reached its maximum after deconvolution from the global-AIF. Further details of the T1-bookend method and processing with the 2 AIFs are available in the Supplemental Data.

### **DWI Analysis**

DWI core infarct was defined, as previously reported,<sup>27</sup> as MD < 0.00057 from DTI mean diffusivity and converted to binary infarction maps. Final infarct volume was calculated from the DTI, which was taken 4 hours post-MCAO, as infarct growth had likely steadied by that time.<sup>16,25</sup>

### **IVIM Analysis**

IVIM local-qCBF (in mL/100 g/min) and diffusion (in mm<sup>2</sup>/s) values were calculated from the 10 b-values by using a 2-step segmented fit.<sup>23,35-39</sup> IVIM D (tissue diffusion) was fit on a voxelwise basis to the second component of the standard IVIM bi-exponential  $\frac{S_b}{S_0} = fe^{-bD^*} + (1-f)e^{-bD}$  for b-values > 250 seconds/mm<sup>2</sup>. After fitting IVIM D, the pseudo diffusion components  $f$  and  $D^*$  were fit to the whole equation by using nonlinear least squares. The parameter  $fD^*$  was quantified as local-qCBF (in mL/100 g/min) by using the water transport time (WTT) model,<sup>24</sup> local-qCBF  $\approx fD^* \times 93,000$  [mL/100 g/min]. IVIM core infarct was defined as IVIM D < 0.000515. This threshold was chosen to agree most closely with DWI total core infarct volume by using leave-one-out cross-validation from the previous study.<sup>40</sup> Further details on IVIM postprocessing are available in the Supplemental Data.

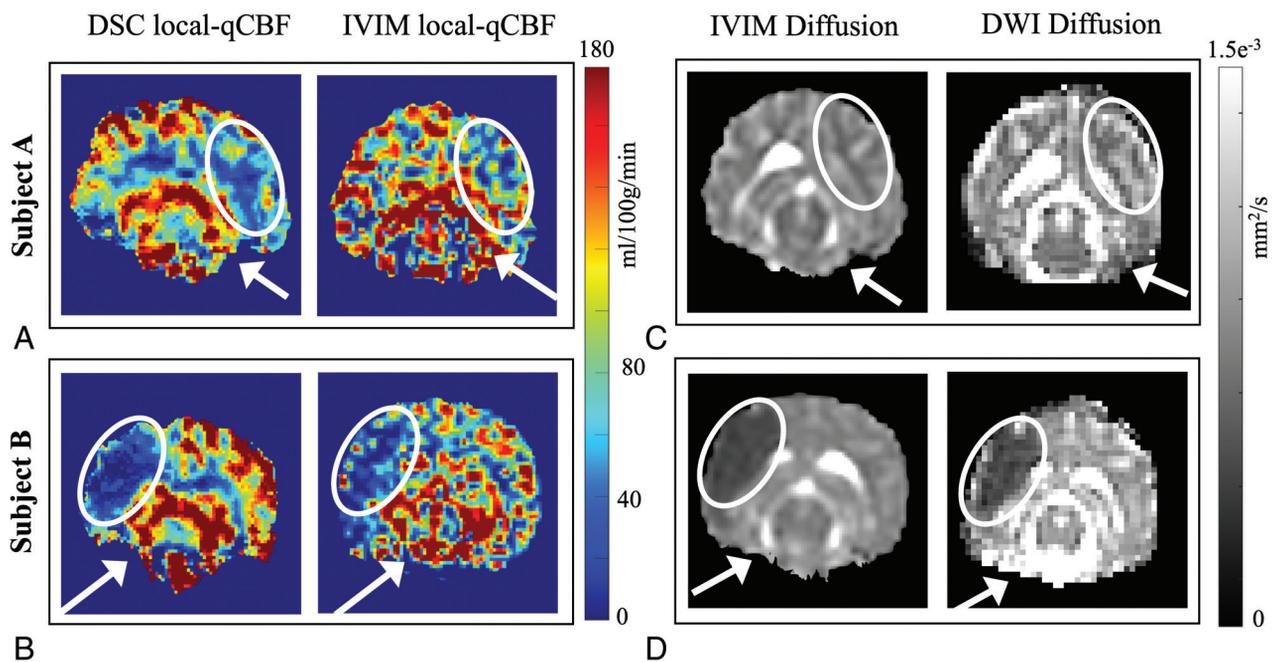
After postprocessing, the IVIM parameter maps were coregistered and resized to overlap anatomically with DSC images for direct region-of-interest comparison. Three consecutive IVIM slices were averaged to match the coarser DSC slice thickness of 6 mm. Previously reported leave-one-out cross-validation T2 map thresholds,<sup>24,40</sup>  $D^* > 0.10$  seconds/mm<sup>2</sup> and  $f > 0.30$ ,<sup>19,41</sup> were applied to remove CSF-dominated voxels and minimize partial volume effects without the use of a T2-prepared IR pulse for CSF suppression.<sup>37</sup>

### **qCBF in Ipsilateral MCA Territory and Final Infarct**

Two and one-half hours postocclusion, the ipsilateral MCA territory was defined as the region of the brain that would normally be supplied by the occluded (right or left) MCA. The average values of IVIM local-qCBF and DSC local-qCBF in the ipsilateral MCA territory that was not infarcted (ie, MD > 0.00057) were compared with each other via linear regression and Bland-Altman. The average DSC and IVIM local-qCBF in this noninfarcted MCA territory were also used to predict final infarct volume by linear regression. IVIM local-qCBF was also compared with DSC standard-qCBF. All analyses were performed in 3 consecutive 6-mm coronal slices starting at, and posterior to, the M1 segment for both DSC and IVIM images.

### **qCBF in the Contralateral MCA Territory**

The contralateral MCA territory was defined as the region of the brain that was supplied by the not occluded (right or left) MCA. The average values of IVIM local-qCBF and DSC local-qCBF in this contralateral MCA territory were compared with each other



**FIG 1.** A comparison of local-qCBF in mL/100 g/min between DSC and IVIM for subject A with good collateral circulation (A) and subject B with poor collateral circulation (B). DSC was fully quantitative and corrected for arterial delay and dispersion effects, and IVIM was quantified with water transport time. The corresponding diffusion maps are shown in (C) and (D), respectively. White arrows denote the position of the MCA coil; white ovals denote the MCA territory. IVIM local-qCBF is averaged across 3 slices to match the 6-mm slice thickness of DSC, and along with an automatic T2 threshold, voxels with  $D^* > 10$  and/or  $f > 0.30$  are excluded to remove fast-flowing CSF-dominated voxels.

via linear regression. IVIM local-qCBF was also compared with DSC standard-qCBF. These analyses were performed in the same 3 slices as the ipsilateral MCA territory.

#### Early Hypoperfused Lesion and Final Infarct Volume

The 2.5 hours hypoperfused lesion was calculated first as the volume of the brain with DSC local-qCBF  $< 26$  mL/100 g/min, second as DSC Tmax  $> 1$  second,<sup>18</sup> and third as IVIM local-qCBF  $< 26$  mL/100 g/min. Correlation of the hypoperfused lesion against the final 4 hours of infarct volume was measured by linear regression. Note that the hypoperfused lesion was independent of DWI measurements; it was only based on thresholded local-qCBF, which could possibly include tissue that had already infarcted. Hypoperfused lesion from standard-qCBF was also calculated by thresholding at 26 mL/100 g/min.

#### Perfusion-Diffusion Mismatch Ratio and Infarct Growth

The PWI/DWI volume ratio was calculated from DSC local-qCBF, IVIM local-qCBF, and Tmax values. The PWI volume was defined as the hypoperfused lesion described above. For DSC local-qCBF, standard-qCBF, and Tmax, the DWI volume was the core infarct ( $MD < 0.00057$  mm<sup>2</sup>/s) volume. For IVIM, the DWI volume was IVIM  $D < 0.000515$  mm<sup>2</sup>/s. Using IVIM D meant that the IVIM PWI/DWI could use images only from the IVIM sequence and allowed simultaneous PWI and DWI. The PWI/DWI ratio of the hypoperfused lesion volume to the core infarct volume was compared via linear regression to the change in infarct volume between the time of the perfusion scans and the final 4 hours infarct. The PWI/DWI mismatch was also calculated from the standard-qCBF hypoperfused lesion.

#### Hypoperfusion from Local Temporal Parameters

Hypoperfusion lesions calculated from slowed transit time from local-AIF DSC MTT and IVIM WTT were generated based on varying transit time thresholds from 1–10 seconds. The volumes from the thresholds were compared to examine correlation and agreement. To reduce the sensitivity to noise, values over 20 seconds were excluded as being “unphysical.” More details on the DSC MTT and IVIM WTT are available in the Supplemental Data.

#### Statistics

Linear regression, Bland-Altman, and Lin Concordance Correlation Coefficient were used to compare IVIM local-qCBF against DSC local-qCBF and standard-qCBF in the MCA territory. Linear regression was used for correlation of local-qCBF and standard-qCBF hypoperfused lesions against DWI final infarct and for correlation of the 2.5 hours PWI/DWI ratio against infarct growth. Paired Wilcoxon signed-rank and linear regression were used to compare all equivalent measures between IVIM, Local-AIF DSC, and standard DSC. All statistical analysis was performed in Python 3.11.4 (Anaconda, 2024). A  $P$  value  $< .05$  was considered statistically significant.

#### RESULTS

Eighty-three percent of the parent study experiments were successful and performed to completion (this current study analyzed the successful cases with completed IVIM, T1-bookend DSC, 2.5 hours, and 4 hours DTI). Access to all results, raw images, and tabulated data are available upon request to the corresponding author.

### Perfusion and Diffusion Maps

Example images of DSC local-qCBF, IVIM local-qCBF maps, IVIM D, and MD infarct maps acquired 2.5 hours post-MCAO are shown in Fig 1, with 3 IVIM slices averaged and resized to match DSC. Subject A was in the control group and had a good native collateral score, as shown by x-ray angiography. The bottom row shows the corresponding images for Subject B in the control group with a poor native collateral score. Note the pronounced difference in ipsilateral perfusion in a setting of good collaterals (Fig 1A versus Fig 1B). In both cases, the IVIM local-qCBF is similar to the DSC local-qCBF. The corresponding maps of the IVIM D and DTI MD are also shown (Fig 1C, -D). Note that in Subject A, the perfusion is reduced (Fig 1A), but tissue viability is maintained (Fig 1C), whereas in

Subject B, both DSC and IVIM local-qCBF showed compromised CBF (Fig 1B) and a large core infarct (Fig 1D) with minimal salvageable PWI-DWI mismatch.

### qCBF in the Ipsilateral MCA Territory

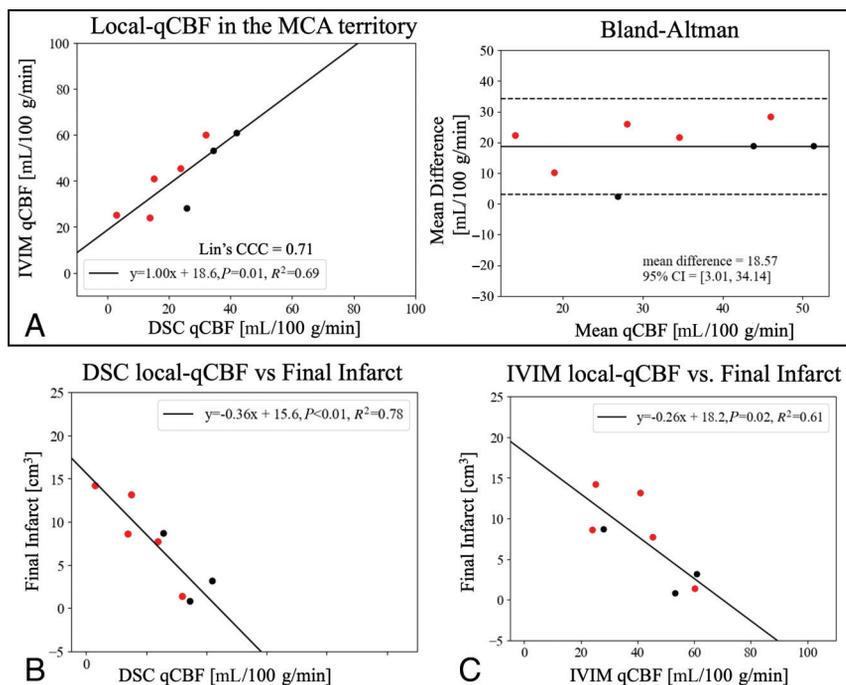
DSC and IVIM local-qCBF in the noninfarcted MCA territory strongly correlated (Fig 2A,  $R^2 = 0.73$ ,  $P = .01$ , Lin CCC = 0.77) across collateral status and flow augmentation. Bland-Altman showed a significant unbiased mean difference (+19 mL/100 g/min), with IVIM local-qCBF higher than DSC local-qCBF (Table). In comparison, the correlation between IVIM local-qCBF and DSC standard-qCBF was not significant ( $R^2 = 0.13$ ,  $P = .38$ ). Lower DSC local-qCBF and IVIM local-qCBF in the noninfarcted ipsilateral MCA territory correlated with larger final infarct ( $R^2 = 0.78$ ,  $0.61$ , respectively; Fig 2B, -C). DSC local-qCBF was the strongest predictor, and IVIM local-qCBF was also a stronger predictor than DSC standard-qCBF for prediction of final infarct ( $R^2 = 0.12$ ).

### qCBF in the Contralateral MCA Territory

Correlation of IVIM local-qCBF in the contralateral MCA territory was similar between DSC local-qCBF and DSC standard-qCBF ( $R^2 = 0.57$ ,  $P = .02$ ;  $R^2 = 0.46$ ,  $P = .05$ , respectively). Again, the IVIM returned higher local-qCBF than both versions of DSC (Table).

### Hypoperfused Lesion Against Final Infarct Volume

A larger 2.5-hour IVIM local-qCBF hypoperfusion lesion predicted a larger final 4-hour infarct volume (Fig 3). IVIM performed similarly to DSC local-qCBF hypoperfusion (Fig 3B, -C) while both outperformed Tmax hypoperfusion (Fig 3A). Local-qCBF hypoperfusion lesions were smaller than the standard-qCBF hypoperfusion lesions (Table), though only DSC local-qCBF was statistically significantly smaller than

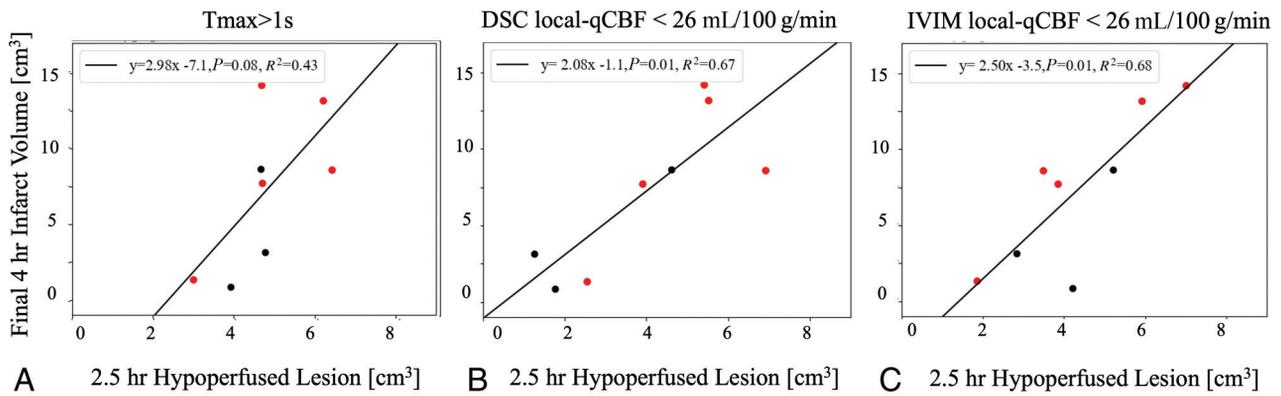


**FIG 2.** A, Correlation and linear regression of IVIM local-qCBF against DSC local-qCBF in the noninfarcted MCA territory 2.5 hours post-MCAO with corresponding Bland-Altman plot. Final (4 hours) infarct volume correlated against (B) DSC local-qCBF and (C) IVIM local-qCBF. Experiments were pooled over treatment and baseline collateralization to expand the range of collateral circulation. Red represents those that received flow augmentation.

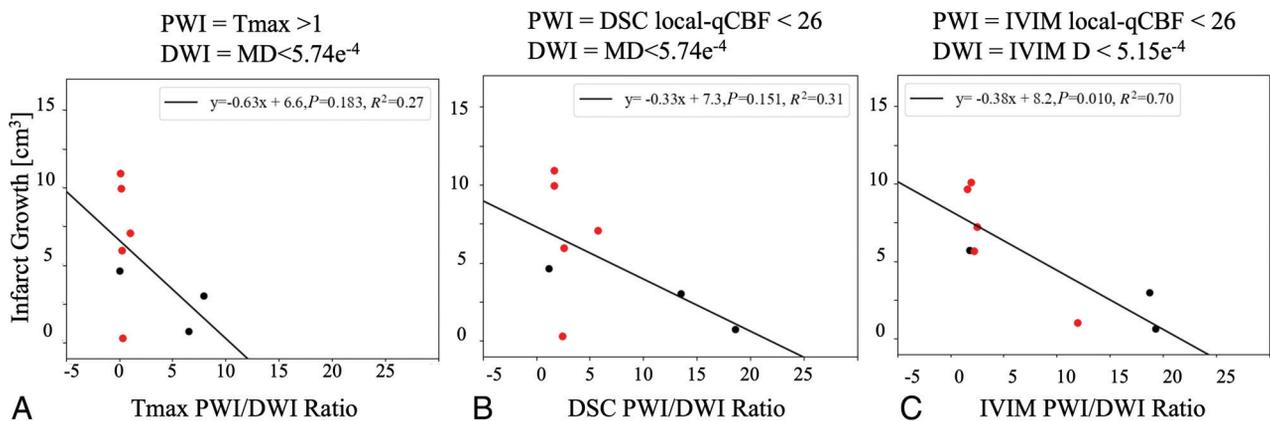
### Comparable parameters between IVIM, local-AIF DSC, and standard DSC<sup>a</sup>

	Local-AIF DSC	IVIM	Standard DSC	IVIM vs Local-AIF DSC	IVIM vs Standard DSC
	$\mu \pm \sigma$ (range)	$\mu \pm \sigma$ (range)	$\mu \pm \sigma$ (range)	Wilcoxon signed-rank: t-stat, P Lin. Regress.: $R^2$ , P	Wilcoxon signed-rank: t-stat, P Lin. Regress.: $R^2$ , P
Noninfarcted ipsilateral MCA territory qCBF	23.9 $\pm$ 11.6 (4.5–44.7)	38.5 $\pm$ 13.5 (34.1–60.8)	17.1 $\pm$ 4.9 (6.8–24.7)	t-stat = 0.0, P = .007 $R^2 = 0.73$ , P = .01	t-stat = 0.0, P = .007 $R^2 = 0.13$ , P = .38
Contralateral MCA territory qCBF	25.1 $\pm$ 14.2 (11.4–47.2)	34.8 $\pm$ 11.9 (14.4–57.1)	14.9 $\pm$ 5.6 (6.2–25.1)	t-stat = 5.0, P = .04 $R^2 = 0.57$ , P = .02	t-stat = 0.0, P = .003 $R^2 = 0.46$ , P = .05
Hypoperfused lesion volume	2.1 $\pm$ 1.1 (0.59–4.2)	2.4 $\pm$ 0.7 (1.5–3.9)	3.4 $\pm$ 1.3 (1.0–5.8)	t-stat = 12.0, P = .46 $R^2 = 0.53$ , P = .18	t-stat = 8.0, P = .19 $R^2 = 0.32$ , P = .44
PWI/DWI ratio	5.9 $\pm$ 6.12 (1.14–18.5)	7.5 $\pm$ 7.4 (1.5–19.3)	12.6 $\pm$ 16.0 (1.3–41.3)	t-stat = 9.0, P = .25 $R^2 = 0.74$ , P = .005	t-stat = 12.0, P = .46 $R^2 = 0.81$ , P = .002

<sup>a</sup> qCBF is presented in mL/100 g/min.



**FIG 3.** The hypoperfused lesion at 2.5 hours post-MCAO was a predictor of final infarct volume for hypoperfusion defined as (A)  $T_{max} > 1s$ , (B) DSC local-qCBF  $< 26$  mL/100 g/min, and (C) IVIM local-qCBF  $< 26$  mL/100 g/min. Experiments were pooled over treatment and baseline collateralization to expand the range of collateral supply. Red represents those that received flow augmentation.



**FIG 4.** The PWI/DWI ratio of the hypoperfused lesion to core infarct at 2.5 hours post-MCAO correlated against the infarct growth between the perfusion scan and the final 4-hour infarct. The PWI lesions are defined by thresholding (A)  $T_{max}$ , (B) DSC local-qCBF, and (C) IVIM local-qCBF. The DWI lesions are defined by (A and B) coregistered mean diffusivity and (C) IVIM D. Experiments were pooled over treatment and baseline collateralization to expand the range of collateral supply. Red represents those that received flow augmentation.

DSC standard-qCBF lesions ( $P = .023$ ). DSC standard-qCBF hypoperfusion lesions showed a weaker correlation with final infarct volume than local-qCBF (slope = 3.24,  $P = .06$ ,  $R^2 = 0.46$  versus Fig 3B, -C).

#### Perfusion-Diffusion Mismatch Ratio and Infarct Growth

The PWI/DWI ratio from  $T_{max}$  and DSC local-qCBF showed a negative correlation with infarct growth (Fig 4A, -B). IVIM PWI/DWI ratio returned a stronger correlation, potentially due to its ability to capture simultaneous perfusion and diffusion in a single scan and avoid coregistration or timing mismatch (Fig 4C).

#### Perfusion Time Parameters

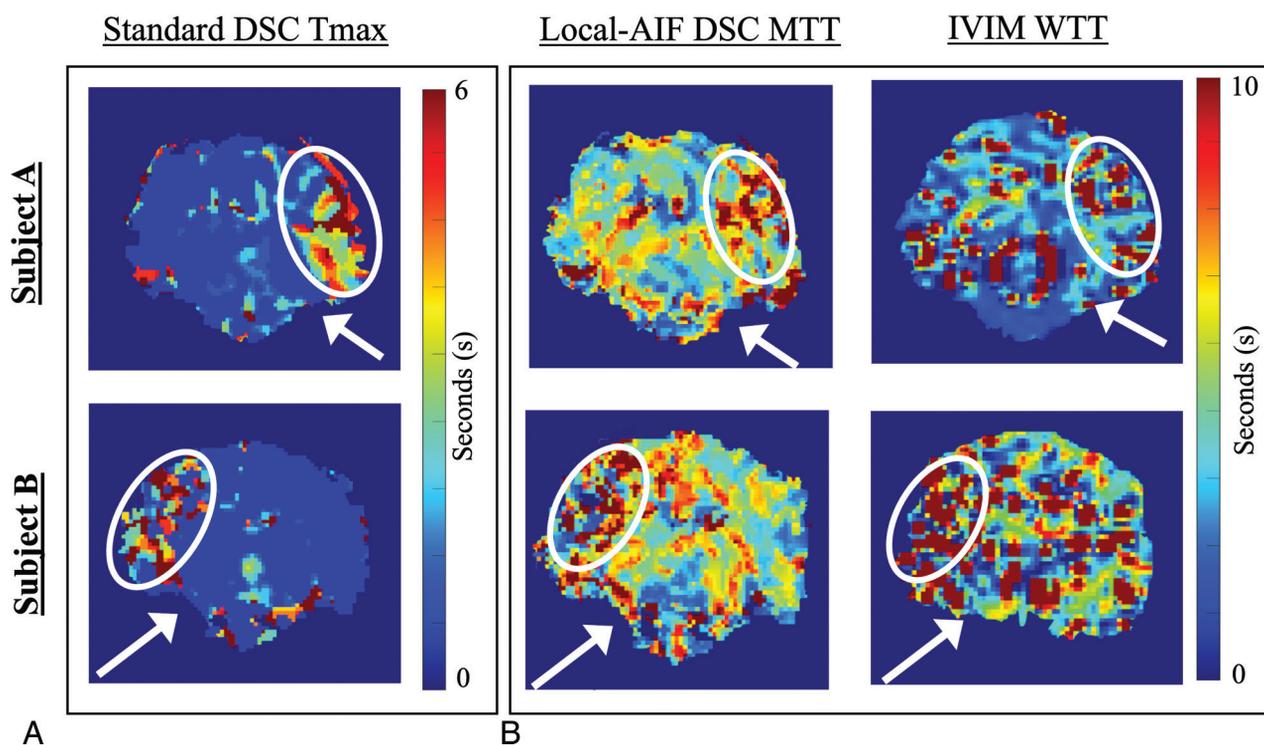
Representative images of DSC  $T_{max}$ , DSC MTT, and IVIM WTT are shown for the subjects from Fig 1 as parametric images in Fig 5. Thresholding IVIM WTT at 3s returned the strongest correlation against DSC MTT at 3 seconds (slope = 0.82,  $P = .04$ ,  $R^2 = 0.60$ ). One case was left out of analysis but included in the plot as a yellow outlier (Supplemental Data). This case highlighted the difficulty of fitting IVIM decay to the standard bi-exponential,<sup>42</sup> as the curve fit returned low IVIM blood fraction  $f$  but normal  $D^*$

supporting the product  $fD^*$  being more accurate than individual parameters due to error correlation.<sup>43</sup>

## DISCUSSION

This study demonstrated IVIM as a method of measuring local-qCBF, including collateral circulation, and perfusion-diffusion mismatch in acute ischemic stroke. DSC and IVIM both independently quantified local-qCBF in mL/100 g/min, correlated strongly, and predicted final infarct volume. As DSC local-qCBF has previously been shown to correlate with collateral score, IVIM local-qCBF correlating strongly with DSC local-qCBF supports IVIM including collateral circulation in acute stroke.<sup>18,19</sup> Further, this work supports the potential benefit of combining temporal parameters and volume into blood flow<sup>44</sup> with both DSC local-qCBF and IVIM local-qCBF outperforming  $T_{max}$  for hypoperfusion and PWI/DWI mismatch. As IVIM is sensitive to local motion of capillary blood, IVIM avoids the complexities of bolus-tracking methods and captures simultaneous quantitative local perfusion and diffusion without contrast agent.

Imaging collateral circulation is valuable to stroke research and therapeutic studies. Even if collateral circulation may arrive delayed and dispersed after having traversed the collateral



**FIG 5.** A, Tmax and B, MTT and IVIM WTT for the same Subject A (*top row, good collaterals*) and Subject B (*bottom row, bad collaterals*) from Fig 1. Tmax is calculated from standard DSC to capture the delayed arterial bolus. MTT is calculated from local-AIF DSC to capture local perfusion speed. WTT is the inverse of  $D^*$  and represents the intravoxel diffusion speed. White arrows denote the position of the MCA coil; white ovals denote the MCA territory. To reduce the fitting sensitivity as the inverse of  $D^*$ , WTT values over 20 seconds were removed from consideration.

network, it can still provide vital nutrients and oxygen to prevent infarction. This may explain why good collateral circulation has been shown to influence endovascular treatment outcome<sup>15</sup> as well as a time-to-treatment window.<sup>45</sup> As such, PWI/DWI mismatch that includes collateral circulation from local-qCBF may help identify patients with good collaterals who can undergo late thrombectomy safely.<sup>5,6,46</sup> Further, research on novel stroke therapeutics that could boost native collateral circulation and extend the stroke treatment window may benefit from an ability to image local-qCBF.<sup>12,13,27,47,48</sup>

As proposed by Liu et al<sup>18</sup> collateral supply can be included in DSC local-qCBF with a local-AIF that corrects delay and dispersion. Further, Federau et al<sup>19</sup> proposed IVIM “local perfusion fraction,” as a measure of collateral blood supply that standard DSC could not capture. The correlation of IVIM local-qCBF to DSC local-qCBF in this current work shows that the 2 methods of measuring “local perfusion” agree. This supports IVIM capturing compensatory native and augmented collateral circulation in acute stroke. In addition, IVIM local-qCBF, DSC local-qCBF, and DSC standard-qCBF agreed in the contralateral MCA territory. This supports the theory that the difference between local-qCBF and standard-qCBF is due to the collateral circulation in the ipsilateral hemisphere that standard-qCBF does not capture.

A subject with higher collateral circulation would be expected to have higher local-qCBF in the ipsilateral MCA territory and smaller final infarct. IVIM local-qCBF demonstrated this trend with higher local-qCBF predicting lower final infarct volume.

The use of flow augmentation to boost collateral circulation in this study meant IVIM could not be directly compared with native collateral score. However, as DSC local-qCBF has demonstrated strong correlation with collateral score in previous works, the agreement of IVIM local-qCBF to DSC local-qCBF supports a similar correlation. It should be noted that the flow augmentation in this study was previously observed to increase collateral circulation only in subjects with lower native collaterals.<sup>13,27</sup> This may be why the treatment group (red) in this work did not always return a higher ipsilateral perfusion with smaller final infarct volume than the control group (black), as seen in Fig 2.

The PWI/DWI mismatch ratio showed negative correlation with future infarct growth. The noninfarcted MCA territory may be receiving collateral circulation with reduced blood flow containing vital nutrients and oxygen. However, there is no guarantee that the collateral circulation will be sustained. Interestingly, the PWI/DWI ratio from IVIM measurements outperformed DSC and Tmax in predicting future infarct growth. As IVIM local-qCBF and DSC local-qCBF themselves showed strong agreement, the improvement of PWI/DWI with IVIM may be due to IVIM imaging simultaneous PWI and DWI, while DSC required a separate DWI MR sequence. Further, IVIM PWI and DWI have the same FOV and resolution, while Tmax and DSC did not match the DWI dimensions, requiring imperfect coregistration.

The positive bias in the Bland-Altman analysis of IVIM perfusion and DSC perfusion (Fig 2B) may be due to motion other than capillary-level blood, such as interstitial fluid and CSF in subarachnoid space. As IVIM is not contrast or spin-labeled, all

motion in a voxel will contribute to signal.<sup>37,38</sup> Since inversion recovery to suppress CSF will also suppress blood,<sup>24,37</sup> instead an automatic T2 threshold was applied to remove CSF-dominated voxels,<sup>24,40</sup> and  $f$  and  $D^*$  were thresholded. However, partial volume contamination could still lead to overestimation of blood signal. Removal with T2-prepared CSF suppression<sup>37</sup> could reduce the offset.

Significant correlation between WTT and MTT supports IVIM WTT as an estimation of local transit time, while the offset highlights limited robustness and the effect of noise in  $D^*$ . Mathematically, the inverse of  $D^*$  may return falsely high WTT values when  $D^*$  is fit to a low value. As the  $D^*$  parameter alone has shown problems with robustness,<sup>49</sup> we only calculated quantitative qCBF as  $fD^*$  averaged across an ROI and for calculation of volumes by thresholding;  $D^*$  was not used as a quantitative value on a voxelwise basis nor for visual analysis.

Previous work has found a correlation between IVIM and DSC<sup>35,38,41</sup> predominately correlating standard DSC CBV and IVIM  $f$ , which avoids the complications of  $D^*$  and may be more reliable. However, the correlations seen in this study by using IVIM local-qCBF and WTT against DSC local-qCBF and MTT show improved correlation, supporting the value of a time component from IVIM and use of a more stable  $D^*$  estimation. Further, IVIM local-qCBF and PWI/DWI ratio in acute stroke has not previously been compared with DSC local-qCBF for prediction of final infarct.

As IVIM is noncontrast, it could capture PWI/DWI mismatch throughout infarct progression without issues of multiple contrast injections.<sup>32</sup> The ability to image longitudinal development of potential penumbra, track perfusion-diffusion mismatch over time, and study infarct growth could aid in preclinical studies of novel stroke therapeutics. One recent study also supports IVIM in the ischemic penumbra correlating with clinical outcome.<sup>50</sup>

This study is not without its limitations. IVIM is still subject to CSF and interstitial fluid contamination, despite use of a T2-weighted threshold. Use of T2 preparation IR pulse<sup>37</sup> may be of benefit in future studies. Gd-contrast prevented comparison of IVIM and DSC throughout infarct progression; if Gd is injected multiple times over the development of infarct progression, some contrast will remain and confound quantification of blood flow. An analysis of IVIM local-qCBF and PWI/DWI mismatch as the stroke develops over time would be a worthwhile investigation. Difficulties related to fitting IVIM data and noise prevented voxelwise comparison to DSC MRI; images were analyzed region-by-region rather than voxel-by-voxel. Further, this work does not study the influence of b-value selection. The complexities of the model and physiologic monitoring prevented perfect temporal agreement of DTI, DSC, and IVIM data. Voxel size and FOV differences prevented perfect coregistration. While heart rate and blood pressure were monitored, fluctuations were inevitable. The conservative number of subjects used in this study limits statistical interpretation and significance. Translation of conclusions derived from animal-based models is a potential limitation, but animal models reduce errors when evaluating methods for qCBF calculation with known occlusion time infarct development which is not possible in humans. While a method of imaging PWI/DWI with collateral circulation throughout infarct development

has use in preclinical research and stroke therapeutics, use of MRI in stroke triage in the United States is limited.

## CONCLUSIONS

IVIM local-qCBF correlated strongly with DSC local-qCBF, hypoperfusion lesion from DSC local-qCBF and IVIM local-qCBF predicted final infarct volume, and IVIM simultaneous PWI/DWI ratio for ischemic penumbra predicted infarct growth. This supports 1) IVIM as a noncontrast method of local-qCBF that includes collateral circulation, 2) improved agreement of IVIM and DSC with a local-AIF, and 3) IVIM as a viable candidate for longitudinal measurement of simultaneous perfusion and diffusion in preclinical stroke research.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Rehani B, Ammanuel SG, Zhang Y, et al. **A new era of extended time window acute stroke interventions guided by imaging.** *Neurohospitalist* 2020;10:29–37 CrossRef Medline
2. Fransen PSS, Beumer D, Berkhemer OA; MR CLEAN Investigators, et al. **MR CLEAN, a multicenter randomized clinical trial of endovascular treatment for acute ischemic stroke in the Netherlands: study protocol for a randomized controlled trial.** *Trials* 2014;15:343 CrossRef Medline
3. Goyal M, Jadhav AP, Bonafe A; SWIFT PRIME Investigators, et al. **Analysis of workflow and time to treatment and the effects on outcome in endovascular treatment of acute ischemic stroke: results from the SWIFT PRIME randomized controlled trial.** *Radiology* 2016;279:888–97 CrossRef Medline
4. Nogueira RG, Jadhav AP, Haussen DC; DAWN Trial Investigators, et al. **Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct.** *N Engl J Med* 2018;378:11–21 CrossRef Medline
5. Olivot J-M, Mlynash M, Thijs VN, et al. **Relationships between cerebral perfusion and reversibility of acute diffusion lesions in DEFUSE.** *Stroke* 2009;40:1692–97 CrossRef Medline
6. Albers GW, Marks MP, Kemp S; DEFUSE 3 Investigators, et al. **Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging.** *N Engl J Med* 2018;378:708–18 CrossRef Medline
7. Lansberg MG, Cereda CW, Mlynash M; Diffusion and Perfusion Imaging Evaluation for Understanding Stroke Evolution 2 (DEFUSE 2) Study Investigators, et al. **Response to endovascular reperfusion is not time-dependent in patients with salvageable tissue.** *Neurology* 2015;85:708–14 CrossRef Medline
8. Heiss W-D, Zaro-Weber O. **Extension of therapeutic window in ischemic stroke by selective mismatch imaging.** *Int J Stroke* 2019; 14:351–58 CrossRef Medline
9. Leigh R, Knutsson L, Zhou J, et al. **Imaging the physiological evolution of the ischemic penumbra in acute ischemic stroke.** *J Cereb Blood Flow Metab* 2018;38:1500–16 CrossRef Medline
10. Demeestere J, Wouters A, Christensen S, et al. **Review of perfusion imaging in acute ischemic stroke.** *Stroke* 2020;51:1017–24 CrossRef Medline
11. Christoforidis GA, Saadat N, Liu M, et al. **Effect of early sanguinate (PEGylated carboxyhemoglobin bovine) infusion on cerebral blood flow to the ischemic core in experimental middle cerebral artery occlusion.** *J Neurointerv Surg* 2022;14:1253–57 CrossRef Medline
12. Shazeeb MS, King RM, Anagnostakou V, et al. **Novel oxygen carrier slows infarct growth in large vessel occlusion dog model based on magnetic resonance imaging analysis.** *Stroke* 2022;53:1363–72 CrossRef Medline

13. Liu M, Saadat N, Jeong YI, et al. **Augmentation of perfusion with simultaneous vasodilator and inotropic agents in experimental acute middle cerebral artery occlusion: a pilot study.** *J Neurointerv Surg* 2023;15:e69–75 CrossRef Medline
14. Lansberg MG, Thijs VN, Hamilton S; DEFUSE Investigators, et al. **Evaluation of the clinical-diffusion and perfusion-diffusion mismatch models in DEFUSE.** *Stroke* 2007;38:1826–30 CrossRef Medline
15. Ribo M, Flores A, Rubiera M, et al. **Extending the time window for endovascular procedures according to collateral pial circulation.** *Stroke* 2011;42:3465–69 CrossRef Medline
16. Christoforidis GA, Vakil P, Ansari SA, et al. **Impact of pial collaterals on infarct growth rate in experimental acute ischemic stroke.** *AJNR Am J Neuroradiol* 2017;38:270–75 CrossRef Medline
17. Maguida G, Shuaib A. **Collateral circulation in ischemic stroke: an updated review.** *J Stroke* 2023;25:179–98 CrossRef Medline
18. Liu MM, Saadat N, Roth SP, et al. **Quantification of collateral supply with local-AIF dynamic susceptibility contrast MRI predicts infarct growth.** *AJNR Am J Neuroradiol* 2025;46:251–58 CrossRef Medline
19. Federau C, Wintermark M, Christensen S, et al. **Collateral blood flow measurement with intravoxel incoherent motion perfusion imaging in hyperacute brain stroke.** *Neurology* 2019;92:e2462–71 CrossRef Medline
20. Wu O, Østergaard L, Koroshetz WJ, et al. **Effects of tracer arrival time on flow estimates in MR perfusion-weighted imaging.** *Magn Reson Med* 2003;50:856–64 CrossRef Medline
21. Jeong YI, Christoforidis GA, Saadat N, et al. **Absolute quantitative MR perfusion and comparison against stable-isotope microspheres.** *Magn Reson Med* 2019;81:3567–77
22. Mouannes-Srour JJ, Shin W, Ansari SA, et al. **Correction for arterial-tissue delay and dispersion in absolute quantitative cerebral perfusion DSC MR imaging.** *Magn Reson Med* 2012;68:495–506 CrossRef Medline
23. Le Bihan D, Breton E, Lallemand D, et al. **Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging.** *Radiology* 1988;168:497–505 CrossRef Medline
24. Liu M, Saadat N, Jeong Y, et al. **Quantitative perfusion and water transport time model from multi b-value diffusion magnetic resonance imaging validated against neutron capture microspheres.** *J Med Imag (Bellingham)* 2023;10:063501 CrossRef
25. Christoforidis GA, Rink C, Kontzialis MS, et al. **An endovascular canine middle cerebral artery occlusion model for the study of leptomeningeal collateral recruitment.** *Invest Radiology* 2011;46:34–40 CrossRef Medline
26. Christoforidis GA, Mohammad Y, Kehagias D, et al. **Angiographic assessment of pial collaterals as a prognostic indicator following intra-arterial thrombolysis for acute ischemic stroke.** *AJNR Am J Neuroradiol* 2005;26:1789–97 Medline
27. Saadat N, Christoforidis GA, Jeong YI, et al. **Influence of simultaneous pressor and vasodilatory agents on the evolution of infarct growth in experimental acute middle cerebral artery occlusion.** *J Neurointerv Surg* 2021;13:741–45 CrossRef Medline
28. Sakaie KE, Shin W, Curtin KR, et al. **Method for improving the accuracy of quantitative cerebral perfusion imaging.** *J Magn Reson Imaging* 2005;21:512–19 CrossRef Medline
29. Vakil P, Lee JJ, Mouannes-Srour JJ, et al. **Cerebrovascular occlusive disease: quantitative cerebral blood flow using dynamic susceptibility contrast MR imaging correlates with quantitative H2[15O] PET.** *Radiology* 2013;266:879–86 CrossRef Medline
30. Carroll TJ, Horowitz S, Shin W, et al. **Quantification of cerebral perfusion using the “bookend technique”: an evaluation in CNS tumors.** *Magn Reson Imaging* 2008;26:1352–59 CrossRef Medline
31. Carroll TJ, Rowley HA, Haughton VM. **Automatic calculation of the arterial input function for cerebral perfusion imaging with MR imaging.** *Radiology* 2003;227:593–600 CrossRef Medline
32. Shin W, Cashen TA, Horowitz SW, et al. **Quantitative CBV measurement from static T1 changes in tissue and correction for intravascular water exchange.** *Magn Reson Med* 2006;56:138–45 CrossRef Medline
33. Shin W, Horowitz S, Ragin A, et al. **Quantitative cerebral perfusion using dynamic susceptibility contrast MRI: evaluation of reproducibility and age- and gender-dependence with fully automatic image postprocessing algorithm.** *Magn Reson Med* 2007;58:1232–41 CrossRef Medline
34. Shah MK, Shin W, Parikh VS, et al. **Quantitative cerebral MR perfusion imaging: preliminary results in stroke.** *J Magn Reson Imaging* 2010;32:796–802 CrossRef Medline
35. Wirestam R, Borg M, Brockstedt S, et al. **Perfusion-related parameters in intravoxel incoherent motion MR imaging compared with CBV and CBF measured by dynamic susceptibility-contrast MR technique.** *Acta Radiology* 2001;42:123–28 CrossRef Medline
36. Federau C, Maeder P, O’Brien K, et al. **Quantitative measurement of brain perfusion with intravoxel incoherent motion MR imaging.** *Radiology* 2012;265:874–81 CrossRef Medline
37. Federau C, O’Brien K. **Increased brain perfusion contrast with T2-prepared intravoxel incoherent motion (T2prep IVIM) MRI.** *NMR Biomed* 2015;28:9–16 CrossRef Medline
38. Federau C, Sumer S, Becce F, et al. **Intravoxel incoherent motion perfusion imaging in acute stroke: initial clinical experience.** *Neuroradiology* 2014;56:629–35 CrossRef Medline
39. Federau C, O’Brien K, Meuli R, et al. **Measuring brain perfusion with intravoxel incoherent motion (IVIM): initial clinical experience.** *J Magn Reson Imaging* 2014;39:624–32 CrossRef Medline
40. Liu M, Warioba C, Bertini J, et al. **Validation and machine learning of a new method for quantifying CBF with IVIM.** *American Society of Neuroradiology Annual Meeting*; Chicago, IL. April 29–May 3, 2023
41. Zhu G, Federau C, Wintermark M, et al. **Comparison of MRI IVIM and MR perfusion imaging in acute ischemic stroke due to large vessel occlusion.** *Int J Stroke* 2020;15:332–42 CrossRef Medline
42. Shazeeb MS, Sotak CH. **Limitations in biexponential fitting of NMR inversion-recovery curves.** *J Magn Reson* 2017;276:14–21 CrossRef Medline
43. Federau C, O’Brien K, Birbaumer A, et al. **Functional mapping of the human visual cortex with intravoxel incoherent motion MRI.** *Plos One* 2015;10:e0117706. CrossRef Medline
44. Nael K, Doshi A, De Leacy R, et al. **MR perfusion to determine the status of collaterals in patients with acute ischemic stroke: a look beyond time maps.** *AJNR Am J Neuroradiol* 2018;39:219–25 CrossRef Medline
45. Hwang YH, Kang DH, Kim YW, et al. **Impact of time-to-reperfusion on outcome in patients with poor collaterals.** *AJNR Am J Neuroradiol* 2015;36:495–500 CrossRef Medline
46. Kim SJ, Seok JM, Bang OY, et al. **MR mismatch profiles in patients with intracranial atherosclerotic stroke: a comprehensive approach comparing stroke subtypes.** *J Cereb Blood Flow Metab* 2009;29:1138–45 CrossRef Medline
47. Warioba C, Liu M, Penano S, et al. **Efficacy assessment of cerebral perfusion augmentation through functional connectivity in an acute canine stroke model.** *AJNR Am J Neuroradiol* 2024;45:1214–19 CrossRef
48. King RM, Anagnostakou V, Shazeeb MS, et al. **Selective brain cooling with a novel catheter reduces infarct growth after recanalization in a canine large vessel occlusion model.** *Interv Neuroradiol* 2024 Jul 23 [Epub ahead of print] CrossRef
49. Wu WC, Chen YF, Tseng HM, et al. **Caveat of measuring perfusion indexes using intravoxel incoherent motion magnetic resonance imaging in the human brain.** *Eur Radiology* 2015;25:2485–92 CrossRef Medline
50. Zimmermann J, Reolon B, Michels L, et al. **Intravoxel incoherent motion imaging in stroke infarct core and penumbra is related to long-term clinical outcome.** *Sci Rep* 2024;14:29631 CrossRef Medline

# Arterial Spin-Labeling MRI Identifies Abnormal Perfusion Metric at the Gray Matter/CSF Interface in Cerebral Small Vessel Disease

 Abdelkader Mahammedi, Ates Fettahoglu,  Jeremy J. Heit, Joanna M. Wardlaw, and  Greg Zaharchuk



## ABSTRACT

**BACKGROUND AND PURPOSE:** Cerebral small vessel disease (SVD) is a common cause of stroke and cognitive decline. SVD is characterized by white matter hyperintensities (WMH) and dilated perivascular spaces (PVS). While WMH can be associated with reduced CBF and glymphatic clearance, current clinical and radiologic assessments of these associations remain controversial and mostly qualitative. We aim to identify if arterial spin-labeling (ASL)-based CBF differences, particularly in the cortical surface at the GM/CSF interface, correlate with SVD severity.

**MATERIALS AND METHODS:** We performed a retrospective cohort study of healthy controls with normal cognition who underwent a brain MRI as part of our university's Alzheimer Disease Research Center (ADRC) and an  $^{15}\text{O}$ -water PET study database. Our inclusion criteria included patients aged  $>50$  years with no structural brain abnormalities besides SVD with ASL perfusion images. WMH grading was performed by using the Fazekas scale, WMH score, PVS grade, and manually segmented WMH volume. We identified patients with moderate-to-severe SVD and then selected age-matched samples of patients with minimal or no SVD. CBF of the whole brain (WB), GM, WM, and along the GM/CSF interface were calculated. Several perfusion metrics (WB, GM, and WM) as well as a novel perfusion metric, normalized GM/CSF interface (nGCI) perfusion metric, which indirectly reflects the relative ASL signal near the GM-CSF boundary, were evaluated by using receiver operating characteristic and correlation analyses.

**RESULTS:** Thirty-two patients met the inclusion criteria ( $n=11$  moderate-to-severe SVD, mean age  $72 \pm 10$  years, 6 women;  $n = 21$  none-to-minimal SVD, mean age  $70 \pm 10$  years, 12 women). Of the measured perfusion markers, nGCI had the strongest negative correlation with Fazekas score, total WMH volume, PVS grade, and average total SVD score ( $r = -0.68, -0.67, -0.54, -0.54$ , respectively;  $P < .001$ ) as well as the highest area under the receiver operating characteristics curve (0.95, 95% CI: 0.87–1.0) as a predictor of WMH severity.

**CONCLUSIONS:** nGCI, a novel perfusion metric that may capture features of perfusion at the GM-CSF boundary, was strongly correlated with WMH and PVS severity. Further, longitudinal studies are required to determine the potential role of nGCI as a predictive marker of SVD progression.

**ABBREVIATIONS:** ADRC = Alzheimer's Disease Research Center; ASL = arterial spin-labeling; nGCI = normalized GM/CSF interface perfusion metric; pcASL = pseudocontinuous arterial spin-labeling; PVS = dilated perivascular space; ROC = receiver operating characteristics; STRIVE = Standards for Reporting Vascular Changes on Neuroimaging; SVD = small vessel disease; WB = whole brain; WMH = white matter hyperintensity

Cerebral small vessel disease (SVD) is a disorder of cerebral microvessels that is associated with stroke and dementia.<sup>1-3</sup> The STAndards for ReportIng Vascular changes on nEuroimaging (STRIVE) criteria have standardized the nomenclature of SVD

markers, which include white matter hyperintensity (WMH) of vascular origin, lacunar infarcts, prominent perivascular spaces (PVS), microbleeds, superficial siderosis, and brain atrophy.<sup>4</sup> Multiple studies suggest that reduced CBF and glymphatic clearance due to disturbance of blood-to-CSF water flow at the BBB or in draining glymphatic pathways may play a critical role in the pathogenesis of SVD, particularly WMH and PVS.<sup>5-7</sup> However, the association between CBF reduction and SVD progression is poorly understood. While SVD markers, particularly WMH, are almost ubiquitous in brain imaging of older adults and common findings on conventional MRI, associations between imaging markers and the heterogeneous clinical expression of SVD remain weak.<sup>8,9</sup> One possible explanation for this discrepancy is that

Received November 12, 2024; accepted after revision December 30.

From the Department of Radiology (A.M., A.F., J.J.H., G.Z.), Stanford University, Stanford, California; and Centre for Clinical Brain and Neuroimaging Sciences (J.M.W.), University of Edinburgh, Edinburgh, Scotland, United Kingdom.

Please address correspondence to Abdelkader Mahammedi, MD, Neuroradiologist, Department of Radiology, Stanford University, 453 Quarry Rd, Palo Alto, CA 94304; e-mail: abdelkader.mahammedi@stanford.edu

 Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8682>

## SUMMARY

**PREVIOUS LITERATURE:** SVD, a disorder of cerebral microvessels, is a common cause of stroke and dementia. Its radiologic correlates, WMH and PVS, are almost ubiquitous in brain imaging as we age. Despite this, their underlying pathogenesis is poorly understood, and associations between imaging markers and the heterogeneous clinical expression of SVD remain weak. While WMH can be associated with reduced CBF and glymphatic clearance, current clinical and radiologic assessments of these associations remain controversial and mostly qualitative.

**KEY FINDINGS:** A novel quantitative perfusion variable, nGCI, was found to have the strongest negative correlation with Fazekas score, total WMH volume, PVS grade, and average total SVD score ( $r = -0.68, -0.67, -0.54, -0.54$ , respectively;  $P < .001$ ). nGCI had the highest area under the ROC curve (0.95, 95% CI: 0.87–1.0) as a predictor of WMH severity.

**KNOWLEDGE ADVANCEMENT:** Decreased ASL signal localized to the GM-CSF boundary was strongly correlated with WMH and PVS severity in elderly subjects with normal cognition.

conventional MRI primarily captures local changes, whereas SVD is now increasingly recognized as a global brain condition that is insufficiently characterized by conventional MRI alone.<sup>1,10,11</sup> Furthermore, the severity and nature of clinical symptoms can be vastly different in patients with seemingly radiologically identical SVD lesions.<sup>12</sup> Therefore, a global approach, such as perfusion imaging with arterial spin-labeling (ASL) MRI, may identify associations between CBF and SVD.

While advances in glymphatic imaging and reduced CBF on ASL have recently gained traction as potential biomarkers for SVD, associations between how SVD may occur and evolve in the setting of prolonged CBF reductions in WM and GM remain unclear and controversial.<sup>5-7,10,13</sup> This debate acknowledges the potential ASL-CBF sampling errors and limited spatial resolution when comparing perfusion between pure GM and WM.<sup>5,6,13-15</sup> Furthermore, prior studies have largely focused on assessing parenchymal ASL signal only, but it is possible to detect ASL signal in CSF spaces as well, though this can be difficult to separate from the adjacent GM.<sup>15,16</sup> Measurement of ASL signal at the GM/CSF interface may enable interrogation of the blood-CSF barrier, which might predict SVD severity and progression.<sup>15</sup> However, no imaging studies have investigated the potential association between the ASL signal at the GM-CSF interface and the severity of SVD. The cortical/superficial perforating arteries primarily provide the arterial supply to the GM, whereas the WM is mainly supplied by the long perforating arteries that originate from cortical arteries and penetrate the subcortical WM.<sup>17,18</sup> While the pathogenesis of SVD is largely unknown, the pathologic processes leading to the arteriolar disease in SVD are mainly believed to result from occlusive disease in the cortical/superficial and deep perforating arteries, which in turn result in decreased and delayed blood flow in the upstream capillaries that supply the WM.<sup>2,11,17,18</sup>

We performed this study to determine if ASL-based CBF measurements, particularly at the GM/CSF interface, correlate with SVD severity. We hypothesized that dysfunction of the cortical/superficial perforating arteries along the GM/CSF interface in patients with high-grade SVD may result in abnormally delayed flow in the capillaries supplying the deep WM in patients with high-grade SVD. Therefore, a lower ASL signal along the GM/CSF interface due to delayed blood flow might be a potential new marker that correlates with a worse WMH and PVS burden. To this end, we isolated the ASL CBF signal from this region and compared it with

other more standard perfusion metrics in cognitively intact elderly subjects with and without moderate-to-severe SVD.

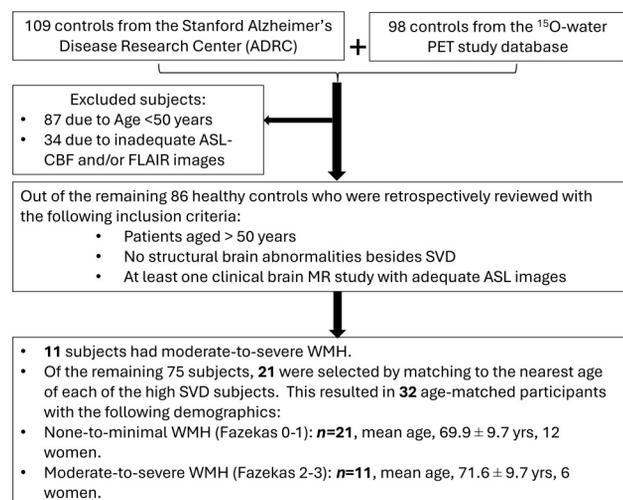
## MATERIALS AND METHODS

### Study Design and Subjects

We performed a retrospective cohort study of consecutive elderly cognitively normal subjects who underwent brain MRI as part of 2 separate local studies: our university's Alzheimer's Disease Research Center (ADRC) and an <sup>15</sup>O-water PET/MR cohort. We included subjects aged >50 years with no structural brain abnormalities besides SVD, who received a brain MR study with adequate quality ASL, T2-weighted, GRE, and T2 FLAIR imaging. Patients were asked to abstain from caffeine for 6 hours, but otherwise, there was no control for potentially confounding variables, such as hematocrit, blood pressure, serum albumin, or time of day. These subjects were analyzed for SVD as described below and further age-matched to create a balanced data set of subjects with and without at least moderate SVD. The study flowchart is shown in Fig 1.

### Image Acquisition, SVD Grading, and ASL-CBF Quantification

All examinations were performed at 3T (Signa PET/MR; GE Healthcare). One neuroradiologist (with 14 years of experience)



**FIG 1.** Data flowchart.

## Patient demographics and imaging characteristics

		Minimal or No WMH (n = 21)	Moderate-to-Severe WMH (n = 11)	All Patients (n = 32)	P Value
Sex					.218
Female		12	6	18	
Male		9	5	14	
Age, Mean ± SD		69.9 ± 9.7	71.6 ± 9.7	70.7 ± 8.9	.636
Imaging characteristics:	MR sequence				
WM disease:	FLAIR				
Fazekas scale (0–3)		[0–1]	[2–3]	[0–3]	<.001
Total WM disease volume (cm <sup>3</sup> )		1.1 ± 0.8	8.3 ± 3.9	4.1 ± 3.4	<.001
Other MR SVD markers:					
Prominent perivascular spaces:	T2WI				.008
None		6	0	6	
1–10		10	1	11	
11–20		4	3	7	
21–40		1	4	5	
>40		0	3	3	
Number of subjects with lacunar infarcts	T2WI/FLAIR	0	2	2	.169
Number of subjects with cerebral microbleeds	T2*GRE/SWI	1	5	6	.054
Avg. total SVD score (0–4)		0.57 ± 0.6	2.3 ± 1.2	1.27 ± 1.2	.002
ASL CBF (mL/100 g/min):	ASL				
Total WM CBF		35.2 ± 8.7	27.5 ± 6.1	31.3 ± 8.4	.036
Total GM CBF		43.5 ± 10.1	36.5 ± 9.1	39.5 ± 9.2	.134
WB CBF		39.1 ± 9.6	33.9 ± 8.3	35.9 ± 8.7	.337
GM/CSF interface CBF		45.8 ± 10.9	36.3 ± 9.1	41.1 ± 9.5	.019
nGCI		1.17 ± 0.04	1.07 ± 0.02	1.12 ± 0.02	<.001

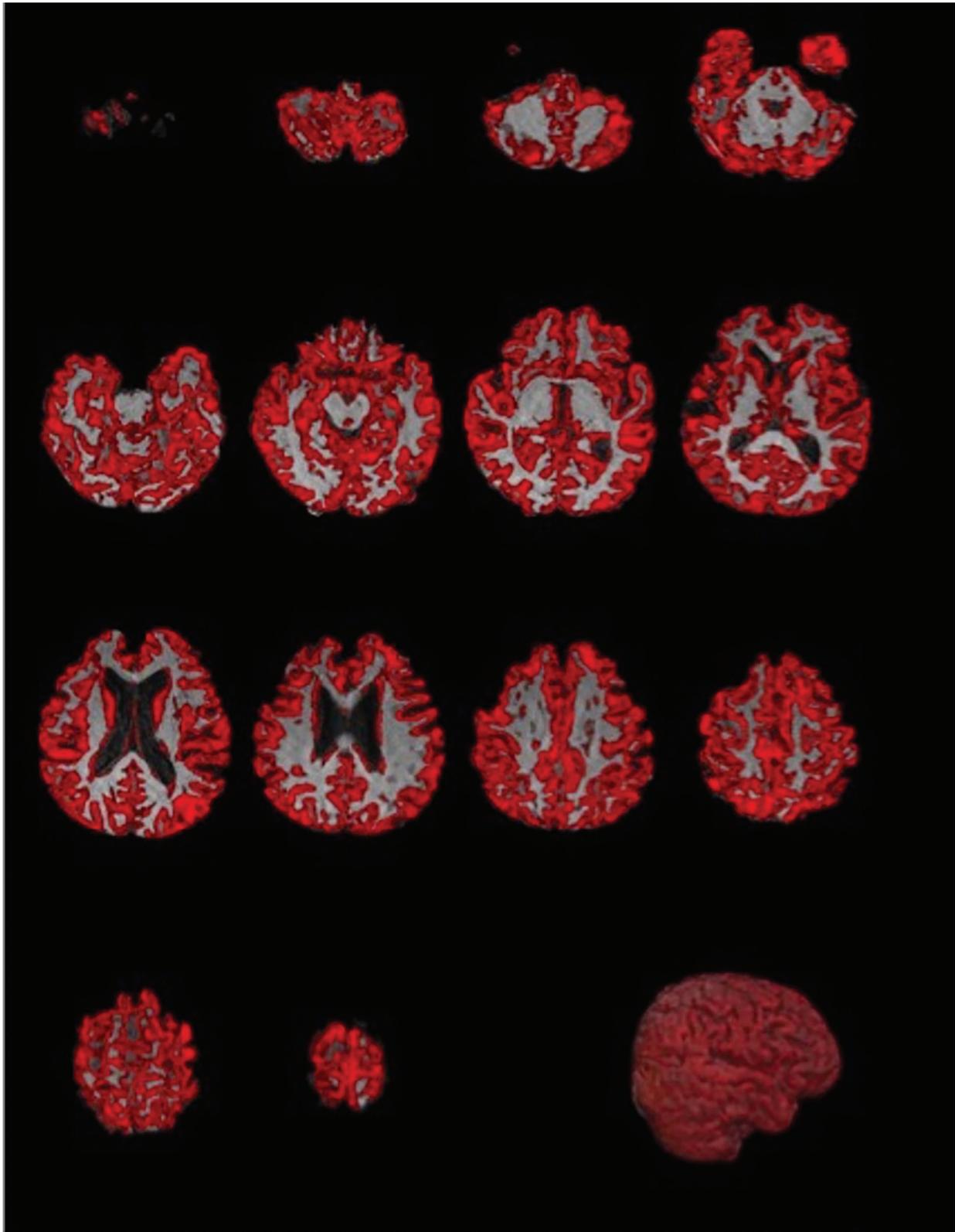
graded the severity of SVD by using the Fazekas score and the STRIVE criteria by using markers of chronic SVD (WMH, PVS, lacunae, and microbleeds) with a total possible SVD score between 0–4.<sup>1,4,19</sup> WMH, PVS, and microbleeds were assessed on the FLAIR, T2-weighted, and GRE images, respectively (Table). Additionally, the total WMH volume was manually segmented by a neuroradiologist and another researcher with 14 and 3 years of experience, respectively. These assessments were performed blinded to the ASL CBF results. A high-resolution 3D gradient-echo T1-weighted structural image was acquired to facilitate image registration with ASL. Parameters for the structural images are reported in the Supplemental Data.

Single-delay 3D segmented stack-of-spirals pseudocontinuous ASL (pcASL) with the following parameters was acquired: labeling duration = 1450 ms, postlabel delay = 2025 ms, TR = 4854 ms, TE = 10.7 ms, arms/points per arm = 8/512, field-of-view = 24 cm, slice thickness = 4 mm, in-plane resolution = 3 mm, number of averages = 3. CBF maps were automatically generated from the ASL difference images by the scanner software following white paper recommendations.<sup>20</sup> These maps were generated at the time of acquisition from pcASL images by using the scanner's reconstruction pipeline. Labeled and control images were reconstructed, and the magnetization difference between labeled and control images was calculated, reflecting the perfusion signal. Using the ASL signal equation,<sup>20</sup> CBF is quantified in units of mL/100 g/min. The labeling duration, postlabel delay, T1 relaxation properties of blood, blood magnetization,

and estimated labeling efficiency are accounted for in this equation to ensure accurate quantification. The resulting CBF maps provide quantitative measures of blood flow across brain tissue and were stored for subsequent analysis.

### Image Processing and Quantitative Analysis

First, ASL-CBF images were registered to the T1-weighted anatomic examination by using FSL's FLIRT (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT>) rigid body registration. T1-weighted images were then processed to create a binarized whole brain (WB) mask by using FSL's brain extraction tool FSL 6.0.5 (Analysis Group, FMRIB). Next, WM and GM segmentation were performed by using FSL's FAST algorithm (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/fast>). GM masks were eroded twice to remove 2 voxel layers from the boundary of nonzero regions by using FSLMATH's default erode function (fslmaths -ero). The resulting eroded GM masks were subtracted from the original GM masks to obtain GM/CSF interface masks. The WM masks were similarly eroded by using the default kernel to prevent overlap with GM voxels. These segmentation masks were overlaid on ASL-CBF images to extract mean WM, GM, and GM/CSF interface CBF, while the binarized brain mask was used to calculate WB CBF. All image analysis was performed in native space for best accuracy in segmenting mask regions, which were uniquely derived for each participant by using T1-structural images. Normalized WM and GM CBF values were also created by dividing the ROI value by the WB CBF. Visual representations of the preprocessing steps are given in Figs 2 and 3. In addition

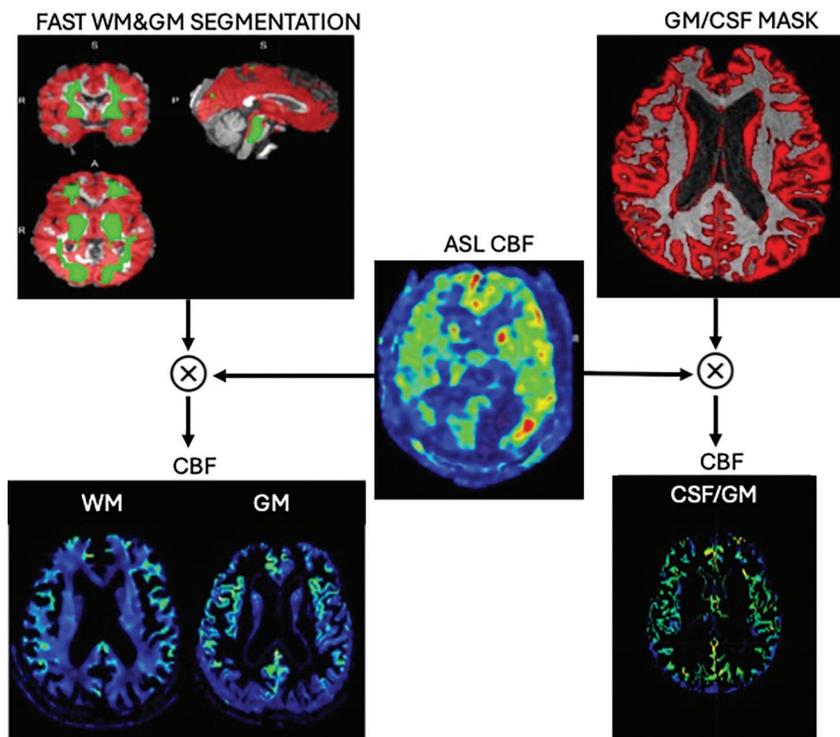


**FIG 2.** GM/CSF interface mask in native space for a representative participant.

to the standard CBF metrics listed above, we normalized the GM/CSF interface ASL-CBF to the WB CBF and have named this parameter normalized GM/CSF interface perfusion metric (nGCI).  $[nGCI = (CBF_{GM/CSF \text{ interface}}) / CBF_{WB}]$

#### **Statistical Analysis**

Continuous variables are presented as mean and standard deviation. Categorical variables are presented as frequencies. Patients were divided into those with none-to-minimal WMH (Fazekas



**FIG 3.** Image processing pipeline for extracting CBF values for GM, WM, WB, and GM/CSF interface from ASL-CBF images. ASL-CBF images are registered to the native T1 structural images. After brain extraction, a binary mask is created from T1 structural images. Segmentation is used to create binary masks for GM and WM. GM/CSF interface masks were obtained through eroding GM masks and subtracting them from the original GM masks. The binary masks are multiplied with the ASL-CBF images in native space to extract the mean CBF value for each masked region.

0–1) versus moderate-to-severe WMH (Fazekas 2–3). The Wilcoxon rank sum test was applied to study the relationship between continuous predictors and WM disease. Fisher exact test was used to evaluate the relationship between categoric predictors with WM disease. The significance threshold was set at a 2-sided  $P$  value  $< .017$  with Benjamini and Hochberg correction for multiple comparisons. The area under the receiver operating characteristics (ROC) curve for separating the cohort into no-to-minimal versus moderate-to-severe SVD was calculated. Comparisons of ROCs for ASL CBF in the different regions were done by using the DeLong method. All statistical analyses were performed by using STATA software (Stata Statistical Software Release 18, StataCorp).

## RESULTS

### Subjects

A total of 207 healthy participants with brain MRI scans were identified from the Stanford ADRC ( $n=109$ ) and the  $^{15}\text{O}$ -water PET study database ( $n=98$ ). Eighty-seven and 34 subjects were excluded because of age  $< 50$  years and technically inadequate ASL-CBF or FLAIR images, respectively. Of the remaining 86 subjects, only 11 had moderate-to-severe WMH based on the Fazekas scale (Table). Then, of the remaining 75 subjects, 21 were selected by matching the sex and nearest age of each of the high SVD subjects. The study flowchart is shown in Fig 1. This resulted in 32 age-matched participants with the following demographics:

none-to-minimal WMH (Fazekas 0–1):  $n=21$ , mean age,  $69.9 \pm 9.7$  years, 12 women; moderate-to-severe WMH (Fazekas 2–3):  $n=11$ , mean age,  $71.6 \pm 9.7$  years, 6 women. The Table highlights the demographic and SVD imaging characteristics within the 2 groups. An example of the imaging data is shown in Fig 4.

### ASL-CBF Quantification

WB, WM, GM, and GM/CSF interface CBF are shown in Fig 4 for the 2 groups. The mean ASL CBF in the WM and along the GM/CSF interface CBF were lower in the group of participants with moderate-to-severe WMH ( $P = .036$  and  $.019$ , respectively; Table).

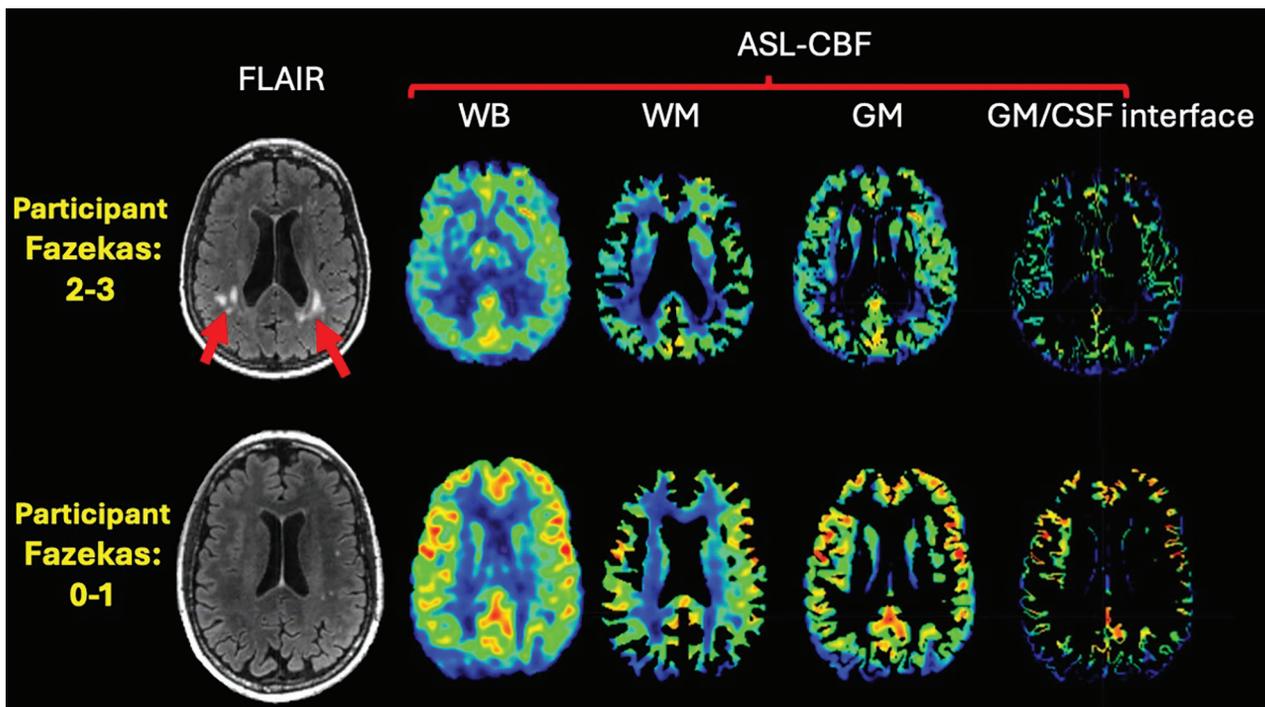
### Relationships between CBF and SVD Metrics

nGCI had the strongest negative correlation with Fazekas score, total WMH volume, PVS grade, and average total SVD score ( $r = -0.68, -0.67, -0.54, -0.54$ , respectively;  $P < .001$ ). While mean ASL measurements in WB, WM, GM, and GM/CSF interface were all lower in the group of participants with higher total SVD score, only WM and GM/CSF interface CBF measurements

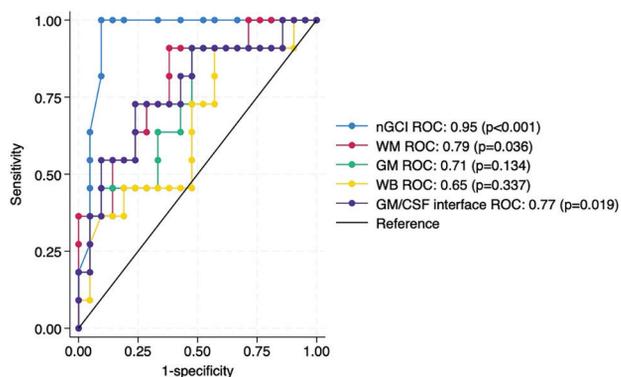
showed a statistically significant difference between the 2 groups ( $P = .036$  and  $0.019$ , respectively), but the highest CBF difference between the 2 groups was along the GM/CSF interface (Table). The mean nGCI was lower in the group of participants with worse SVD burden (Table). The nGCI measurement had the highest area under the ROC curve (0.95, 95% CI: 0.87–1.0) (Fig 5) to separate Fazekas 0–1 from Fazekas 2–3 patients. The total WM, GM, WB, and GM/CSF interface CBF had the following areas under the receiver operating characteristics curve (0.79 [95% CI: 0.62–0.95], 0.71 [95% CI: 0.52–0.91], 0.65 [95% CI: 0.44–0.86], and 0.77 [95% CI: 0.58–0.95], respectively).

## DISCUSSION

In this study, we evaluated ASL-based CBF measurements in different brain regions and correlated them with SVD markers in cognitively normal older individuals. We found that the normalized ASL signal along the GM/CSF interface (nGCI) was lower in those with higher Fazekas scores. We also found that subjects with lower CBF along GM/CSF interface had worse total WMH volume, PVS grade, and average total SVD score. These findings may advance our understanding of the quality of blood flow to the brain in patients with SVD. To our knowledge, no previous study has characterized ASL CBF of the GM/CSF interface. However, consistent with previous studies, reduced global WM CBF in patients with higher grade WMH has been previously demonstrated by using multiple modalities, such as xenon-CT,



**FIG 4.** CBF maps of the WB, WM, GM, and GM/CSF interface by using ASL in 2 subjects with minimal versus moderate-to-severe WMH (Fazekas score: 0–1 versus 2–3).



**FIG 5.** ROC curves for multiple perfusion metrics (WM, GM, WB, GM/CSF interface, and nGCI-CBF) to predict WMH severity (Fazekas 0–1 versus 2–3). nGCI-CBF had the highest area under the ROC curve (0.95).

PET, SPECT, and MRI (including ASL).<sup>5,6,21</sup> While multiple cross-sectional studies showed an association between more WMHs and reduced CBF, longitudinal associations varied and remain unclear.<sup>5,6</sup> Our findings concord with Han et al,<sup>6</sup> who found a decreasing trend of WM and GM CBF with higher WMH burden. Notably, in their study, the association with GM CBF was not significant after adjusting for age and sex.

The arterial supply to the GM and WM of the brain differs in attenuation and distribution due to the distinct metabolic demands of these tissues. While the arterial supply of the GM primarily comes from the cortical and superficial perforating arteries, the WM supply is mainly from the longer perforating arteries, which extend from cortical arteries and enter the subcortical WM.<sup>17,18</sup> These arteries are long and often begin to coil and spiral as they enter the WM. It was suggested that WM CBF

could depend on the proximity to the cortex, given that arteries perfusing the WM are branches originating from the cortical surface, making the WM more susceptible to ischemia due to fewer collateral pathways.<sup>17</sup> Chronic ischemia is characterized by CBF decline and has been identified as a significant contributor to the development of SVD, specifically in the cortical and superficial perforating arteries near the blood-to-CSF barrier at the cortical surface, which are particularly vulnerable to CBF reductions.<sup>2,11,17</sup> We speculate that these differences in ASL signal along the GM/CSF interface between patients with minimal versus high-grade SVD could be due to delayed blood flow and longer than average arrival times in the group with high-grade SVD, though this could not be proved by using a single-delay ASL sequence. This delayed blood flow, which is an index of cortical perfusion, may reflect dysfunction in the cortical/superficial perforating arteries, leading to abnormally delayed flow in the upstream capillaries supplying the deep WM. These disruptions could correspond to regions where WMH and PVS manifest, potentially indicating future sites responsible for SVD-related clinical complications, such as subcortical acute ischemic events and/or vascular cognitive impairment.

Our finding that nGCI correlated with the total SVD score may provide a more practical framework for better evaluating overall brain health. SVD is an increasingly large health burden that is associated with about 25% of ischemic strokes, most hemorrhagic strokes, and vascular dementia.<sup>11,22</sup> Recently, the strict boundaries separating cognitive impairment, vascular dementia, stroke, and SVD have faded, with a greater emphasis on overall brain health, which can be assessed by using SVD MR markers.<sup>8,11,23,24</sup> Although the STRIVE criteria have standardized the nomenclature of SVD markers, studies have shown that

combining these MR SVD markers and using a more comprehensive approach to assess total SVD burden<sup>19</sup> can offer a more pragmatic way of assessing overall brain health, which has been shown to be a prognostic indicator of cognitive impairment, gait and mood disorders, and recurrent stroke.<sup>19,25,26</sup> However, in practice, distinguishing between the different individual neuroimaging markers of SVD can be challenging and time-consuming, as they are often closely interconnected. Therefore, our proposed nGCI may better inform on overall brain health.

Reduced CBF on ASL has recently gained traction as a potential biomarker for SVD that may play a key role in the development of SVD, though associations between SVD and reduced CBF in WM and GM remain controversial, likely secondary to perfusion measurement errors.<sup>5-7,13-15,21</sup> For better quantification of perfusion with ASL, partial volume correction can be used to disentangle the signals from GM and WM within any voxel.<sup>15,16</sup> Traditionally, the CSF contribution to the total ASL perfusion signal is assumed to be zero, and only GM and WM perfusion values are extracted. However, it was shown in a recent study that labeled water does, in fact, exchange into the CSF to create a measurable CSF-ASL signal, which can be visualized by using long TE ASL sequences. Such sequences are not used for standard clinical ASL,<sup>16</sup> including in the current study. However, including such sequences in research studies of SVD may be helpful in better understanding the glymphatic waste clearance system and lead to a better understanding of SVD pathogenesis. Correlation of potential CSF-related ASL signal with measurements that could be made on clinical ASL sequences, such as nGCI, might enable more widespread evaluation of any possible glymphatic component to ASL.

While Petitclerc et al<sup>16</sup> proposed a modified ASL MRI sequence by using ultra-long-TE ASL to extract and directly measure the CSF-ASL signal contribution at the blood-CSF barrier, doing so in our study is difficult since we used standard clinical ASL parameters only. However, our findings may support but cannot prove the CSF-ASL contribution as we demonstrated that CBF changes along the GM/CSF interface have the highest correlation with WMH and PVS severity in elderly subjects with normal cognition, which are best captured by the nGCI perfusion metric, a potential novel marker of SVD severity. Further, longitudinal studies are required to determine the potential role of nGCI as a predictive marker of future SVD progression.

Several limitations of this study should be noted. First, this is a small, retrospective, single-site, exploratory study. Since this was not the point of the original studies, only a few subjects had severe WMH, which limits the evaluation of a wide range of SVD. All subjects were scanned on the same scanner, which probably improved any variability due to vendor, but which may not translate to studies done on other scanner types. Further, the use of an eroded GM mask to probe the GM-CSF boundary may be limited but it was chosen as the best objective way to identify relevant voxels. Additionally, while multidelay ASL provides more accurate CBF quantification by accounting for variations in arterial transit times, particularly in the SVD populations, most of our subjects had only single-delay ASL. Also, the postlabel delay of 2025 ms, while recommended by prior white papers for clinical imaging, may be insufficiently long to capture flow if it is

severely delayed; this is a fundamental challenge with ASL, balancing the sensitivity of the technique with sensitivity to very slow flow. Future studies should include multidelay ASL with long labeling and longer postlabel delays.<sup>27</sup> While age and sex matching reduce confounding, certain vascular risk factors (eg, hypertension, diabetes, smoking) may result in exogenous variation of CBF measurements. However, all the participating subjects were healthy controls without a known medical history of vascular risk factors.

## CONCLUSIONS

We have evaluated ASL-based perfusion in multiple brain regions in patients with and without moderate-to-severe SVD markers. We found a potential marker related to the GM-CSF junction that might be associated with worse SVD. Future, larger, prospective studies, including longitudinal follow-up, could test whether this metric is indeed related to the development of SVD.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Mahammedi A, Wang LL, Williamson BJ, et al. **Small vessel disease, a marker of brain health: what the radiologist needs to know.** *AJNR Am J Neuroradiol* 2022;43:650–60 CrossRef Medline
2. Wardlaw JM, Smith C, Dichgans M. **Mechanisms of sporadic cerebral small vessel disease: insights from neuroimaging.** *Lancet Neurol* 2013;12:483–97 CrossRef Medline
3. Pantoni L. **Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges.** *Lancet Neurol* 2010;9:689–701 CrossRef Medline
4. Wardlaw JM, Smith EE, Biessels GJ; Standards for Reporting Vascular changes on neuroimaging (STRIVE v1), et al. **Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration.** *Lancet Neurol* 2013;12:822–38 CrossRef Medline
5. Stewart CR, Stringer MS, Shi Y, et al. **Associations between white matter hyperintensity burden, cerebral blood flow and transit time in small vessel disease: an updated meta-analysis.** *Front Neurol* 2021;12:647848 CrossRef Medline
6. Han H, Ning Z, Yang D, et al. **Associations between cerebral blood flow and progression of white matter hyperintensity in community-dwelling adults: a longitudinal cohort study.** *Quant Imaging Med Surg* 2022;12:4151–65 CrossRef Medline
7. Mestre H, Kostrikov S, Mehta RI, et al. **Perivascular spaces, glymphatic dysfunction, and small vessel disease.** *Clin Sci (Lond)* 2017; 131:2257–74 CrossRef Medline
8. Shibuya M, Leite C da C, Lucato LT. **Neuroimaging in cerebral small vessel disease: update and new concepts.** *Dement Neuropsychol* 2017;11:336–42 CrossRef Medline
9. Gouw AA, Seewann A, van der Flier WM, et al. **Heterogeneity of small vessel disease: a systematic review of MRI and histopathology correlations.** *J Neurol Neurosurg Psychiatry* 2011;82:126–35 CrossRef Medline
10. Shi Y, Thrippleton MJ, Blair GW, et al. **Small vessel disease is associated with altered cerebrovascular pulsatility but not resting cerebral blood flow.** *J Cereb Blood Flow Metab* 2020;40:85–99 CrossRef Medline
11. Wardlaw JM, Smith C, Dichgans M. **Small vessel disease: mechanisms and clinical implications.** *Lancet Neurol* 2019;18:684–96 CrossRef Medline
12. Ter Telgte A, van Leijsen EMC, Wiegertjes K, et al. **Cerebral small vessel disease: from a focal to a global perspective.** *Nat Rev Neurol* 2018;14:387–98 CrossRef Medline

13. Joseph CR. Utilizing 3D arterial spin labeling to identify cerebrovascular leak and glymphatic obstruction in neurodegenerative disease. *Diagnosics (Basel)* 2021;11:1888 CrossRef Medline
14. Li Y, Li M, Zuo L, et al. Compromised blood–brain barrier integrity is associated with total magnetic resonance imaging burden of cerebral small vessel disease. *Front Neurol* 2018;9:221 CrossRef Medline
15. Petitsclerc L, Hirschler L, Örszik B, et al. Arterial spin labeling signal in the CSF: implications for partial volume correction and blood–CSF barrier characterization. *NMR Biomed* 2023;36:e4852 CrossRef Medline
16. Petitsclerc L, Hirschler L, Wells JA, et al. Ultra-long-TE arterial spin labeling reveals rapid and brain-wide blood-to-CSF water transport in humans. *Neuroimage* 2021;245:118755 CrossRef Medline
17. Aslan S, Huang H, Uh J, et al. White matter cerebral blood flow is inversely correlated with structural and functional connectivity in the human brain. *Neuroimage* 2011;56:1145–53 CrossRef Medline
18. Nonaka H, Akima M, Hatori T, et al. The microvasculature of the cerebral white matter: arteries of the subcortical white matter. *J Neuropathol Exp Neurol* 2003;62:154–61 CrossRef Medline
19. Staals J, Makin SDJ, Doubal FN, et al. Stroke subtype, vascular risk factors, and total MRI brain small-vessel disease burden. *Neurology* 2014;83:1228–34 CrossRef Medline
20. Alsop DC, Detre JA, Golay X, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: a consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn Reson Med* 2015;73:102–16 CrossRef Medline
21. Promjunyakul N, Lahna D, Kaye JA, et al. Characterizing the white matter hyperintensity penumbra with cerebral blood flow measures. *Neuroimage Clin* 2015;8:224–29 CrossRef Medline
22. Debette S, Schilling S, Duperron MG, et al. Clinical significance of magnetic resonance imaging markers of vascular brain injury: a systematic review and meta-analysis. *JAMA Neurol* 2019;76:81–94 CrossRef Medline
23. Kapasi A, DeCarli C, Schneider JA. Impact of multiple pathologies on the threshold for clinically overt dementia. *Acta Neuropathol* 2017;134:171–86 CrossRef Medline
24. Gorelick PB, Furie KL, Iadecola C, et al. Defining optimal brain health in adults. *Stroke* 2017;48:e284–303
25. Banerjee G, Jang H, Kim HJ, et al. Total MRI small vessel disease burden correlates with cognitive performance, cortical atrophy, and network measures in a memory clinic population. *J Alzheimers Dis* 2018;63:1485–97 CrossRef Medline
26. Charidimou A, Martinez-Ramirez S, Reijmer YD, et al. Total magnetic resonance imaging burden of small vessel disease in cerebral amyloid angiopathy: an imaging-pathologic study of concept validation. *JAMA Neurol* 2016;73:994–1001 CrossRef Medline
27. Fan AP, Guo J, Khalighi MM, et al. Long-delay arterial spin labeling provides more accurate cerebral blood flow measurements in moyamoya patients: a simultaneous positron emission tomography/MRI study. *Stroke* 2017;48:2441–49 CrossRef Medline

# Correlation between P2-PCA Volume Flow Rate and BOLD Cerebrovascular Reactivity in Patients with Symptomatic Carotid Artery Occlusion

Amely Walser,  Jorn Fierstra, Lara Maria Hübner,  Jacopo Bellomo,  Tilman Schubert, Menno Germans,  Vittorio Stumpo,  Christiaan Hendrik Bas van Niftrik,  Susanne Wegener, Andreas R. Luft, Zsolt Kulcsár,  Luca Regli, Giuseppe Esposito, and Martina Sebök

## ABSTRACT

**BACKGROUND AND PURPOSE:** Identifying and assessing hemodynamic and flow status in patients with symptomatic ICA occlusion is crucial for evaluating recurrent stroke risk. The aim of this study was to analyze the correlation between 2 quantitative imaging modalities: 1) blood oxygenation level–dependent (BOLD) cerebrovascular reactivity (CVR) and 2) quantitative MR angiography (qMRA) with noninvasive optimal vessel analysis (NOVA), measuring volume flow rate (VFR). Comparing these modalities is relevant for assessing collateral circulation and hemodynamic impairment.

**MATERIALS AND METHODS:** In this retrospective analysis of prospectively collected data, 37 symptomatic patients with unilateral ICA occlusion, who underwent both NOVA-qMRA and BOLD-CVR investigation, were included. The correlation analysis between NOVA-qMRA–derived second segment of the posterior cerebral artery (PCA-P2) VFR and BOLD-CVR (hemispheric and MCA territory CVR) was done by using a linear mixed-effects model.

**RESULTS:** A moderate correlation was found between P2-VFR and BOLD-CVR values for the ipsilateral MCA territory ( $r = 0.44$ ,  $R^2 = 0.2$ ,  $P < .001$ ) and the ipsilateral hemisphere ( $r = 0.39$ ,  $R^2 = 0.15$ ,  $P < .001$ ), indicating that 20% of the variance in P2-VFR can be explained by the BOLD-CVR of the MCA territory and 15% by the BOLD-CVR of the affected hemisphere.

**CONCLUSIONS:** This correlation suggests that impaired BOLD-CVR is partly linked to an increased PCA-P2 volume flow rate, potentially indicating the activation of leptomeningeal collaterals in severe hemodynamic conditions. Both imaging techniques could aid clinicians in creating personalized treatment strategies for patients with symptomatic ICA occlusion.

**ABBREVIATIONS:** ACA = anterior cerebral artery; BOLD = blood oxygenation level–dependent; CVR = cerebrovascular reactivity; M1 = first segment of the middle cerebral artery; NOVA = noninvasive optimal vessel analysis; P2 = second segment of the posterior cerebral artery; PCA = posterior cerebral artery; qMRA = quantitative MRA; TCD = transcranial Doppler; VFR = volume flow rate

Occlusion of the ICA is one of the most common causes of ischemic stroke and is highly clinically relevant due to its association with poor outcomes.<sup>1</sup> Up to 50% of the patients with ICA occlusion do not show any symptoms and are diagnosed incidentally.<sup>2</sup> The clinical outcome of ICA occlusion—whether it remains asymptomatic or progresses to an ischemic stroke—largely depends on the adequacy of collateral flow.<sup>3,4</sup> Identifying

hemodynamic impairment by using advanced quantitative neuroimaging is crucial, as it enables the prediction of future stroke risk and informs tailored management strategies, particularly for patients with symptomatic ICA occlusion. Our recent study demonstrated that patients with symptomatic large vessel cerebrovascular disease and impaired BOLD-CVR in the affected hemisphere had a 10.7-fold higher risk of recurrent ischemic stroke compared with those with preserved BOLD-CVR.<sup>5</sup> The degree of hemodynamic impairment caused by ICA occlusion can be evaluated by using a range of imaging techniques, including the reference standard [<sup>15</sup>O]H<sub>2</sub>O-PET and SPECT techniques with vasodilatory challenge.<sup>6</sup> Because of limited clinical availability (use of exogenous radioisotope), daily use of the above-mentioned imaging techniques in clinical practice is not warranted.<sup>7,8</sup>

In the last decade, 2 novel quantitative and noninvasive MRI techniques have been introduced that allow a more precise hemodynamic and collateral flow assessment: 1) blood oxygenation level–dependent (BOLD) functional MRI with a vasodilatory

Received October 29, 2024; accepted after revision December 14.

From the Departments of Neurosurgery (G.E., M.S., A.W., J.F., L.M.H., J.B., M.G., V.S., C.H.B.v.N., L.R.), Neuroradiology (T.S., Z.K.), Neurology (S.W., A.R.L.), and Clinical Neuroscience Center (M.S., A.W., J.F., L.M.H., J.B., M.G., V.S., C.H.B.v.N., L.R., T.S., S.W., A.R.L., Z.K.), University Hospital Zurich, University of Zurich, Zurich, Switzerland; and Cereneo Center for Neurology and Rehabilitation (A.R.L.), Vitznau, Switzerland.

This research was funded by the Clinical Research Priority Program of the University of Zurich (UZH CRPP Stroke).

Giuseppe Esposito and Martina Sebök contributed equally to this article.

Please address correspondence to Martina Sebök, MD, PhD, Department of Neurosurgery, University Hospital Zurich, Frauenklinikstrasse 10, 8091 Zurich, Switzerland; e-mail: martina.seboek@usz.ch

<http://dx.doi.org/10.3174/ajnr.A8626>

## SUMMARY

**PREVIOUS LITERATURE:** Patients with symptomatic ICA occlusion are at a high risk of recurrent ischemic stroke, particularly when hemodynamic impairment is present. This highlights the need for precise hemodynamic and blood flow assessments. This study evaluates the role of 2 advanced imaging techniques, NOVA-qMRA and BOLD-CVR, focusing on their utility in assessing P2 VFR and CVR. Previous research using transcranial Doppler has demonstrated increased P2-PCA flow velocity as an indicator of leptomeningeal collateral activation under hemodynamic compromise.

**KEY FINDINGS:** A moderate correlation between P2-VFR and BOLD-CVR values in the ipsilateral MCA territory and hemisphere suggests that reduced BOLD-CVR may be partially associated with increased PCA-P2 volume flow rate. This finding potentially indicates the activation of leptomeningeal collaterals under severe hemodynamic compromise.

**KNOWLEDGE ADVANCEMENT:** The combination of BOLD-CVR and NOVA-qMRA enables early detection of hemodynamic impairment and assessment of blood flow compensatory mechanisms, critical factors in the recurrent stroke risk. These advanced imaging techniques can assist in stroke risk stratification and guide the development of personalized treatment approaches, including surgical or endovascular revascularization strategies.

CO<sub>2</sub> challenge to assess cerebrovascular reactivity (CVR) at the brain parenchyma level<sup>5,9</sup> and 2) quantitative MR angiography (qMRA) with noninvasive optimal vessel analysis (NOVA) for measurement of volume flow rate (VFR) in mL/min in large cerebral vessels allowing determination of the status of collateral circulation.<sup>10</sup> VFR informs on collateral status by quantifying the blood flow in major cerebral vessels, thereby indicating the capacity of collateral vessels to compensate for impaired circulation. An activation of collaterals means an increase in blood flow in these vessels as they become “activated” or “predilated” to maintain adequate cerebral perfusion. Recently, a study comparing [<sup>15</sup>O]H<sub>2</sub>O-PET and BOLD-CVR showed that BOLD-CVR examinations harbor sufficient data to provide specific cerebrovascular reactivity cutoff points for hemodynamic failure staging.<sup>5</sup> Moreover, previous studies have emphasized the critical role of ipsilateral leptomeningeal collateral activation<sup>11</sup> and the association with impaired hemodynamic status. In these cases, ipsilateral leptomeningeal collateral activation was identified as the only collateral pathway linked to severely impaired BOLD-CVR.<sup>8</sup>

Both parameters—cerebrovascular reactivity and volume flow rate—can be altered in patients with ICA occlusion; however, the correlation between these 2 parameters remains unknown. The aim of this study is to analyze the correlation between the BOLD-CVR and qMRA-NOVA VFR in patients with symptomatic ICA occlusion. Understanding this correlation is crucial for improving the assessment of hemodynamic status and facilitating guidance on the treatment of patients.

## MATERIALS AND METHODS

This study was approved by the Ethics Committee of the Canton of Zurich with the following number: KEK 2020–02314. All participants signed a general informed consent form before they were included in the study. This study adhered to the methodology outlined in the STROBE checklist.

### Patient Selection

In this retrospective analysis of prospectively collected data, we included patients with symptomatic ICA occlusion who underwent both BOLD-CVR and NOVA-qMRA imaging within a

2-week period at the Clinical Neuroscience Center of the University Hospital Zurich. A portion of the patients were part of “The Interplay of Microcirculation and Plasticity After Ischemic Stroke (IMPreST) Study,” which is part of the Clinical Research Priority Programs Stroke funded by the University of Zurich (ClinicalTrials.gov identifier: NCT04035746). The cause of ICA occlusion in all included patients was atherosclerosis. Patients with ICA occlusion caused by other etiologies, such as trauma or Moyamoya disease, were excluded from the study.

Patients with bilateral ICA stenosis or occlusion and those with asymptomatic, incidental findings of ICA occlusion were excluded from the study to avoid confounding effects from severely reduced collateral flow that could independently influence both BOLD-CVR and qMRA-NOVA VFR measurements. However, patients with contralateral ICA stenosis were included because their hemodynamic status allows for adequate collateral flow.

The demographic data, medical history, and relevant clinical information were reviewed through medical chart reviews under the supervision of stroke neurologists. The contralateral asymptomatic ICA stenosis was categorized into low-grade (<50%), moderate-grade (50%–69%), and high-grade (≥70%) according to the criteria of the North American Symptomatic Carotid Endarterectomy Trial (NASCET).<sup>12</sup>

**Image Acquisition and Analysis.** For each patient, the BOLD-CVR data were collected for the whole brain, the hemispheres, and the anterior cerebral artery (ACA), MCA, and posterior cerebral artery (PCA) territory in both hemispheres (ipsilateral to the occlusion and contralateral hemisphere).<sup>2</sup> With NOVA accordingly, the VFR of the A2 segment of the anterior cerebral artery (A2-ACA), the M1 segment of the middle cerebral artery (M1-MCA), and the P2 segment of the posterior cerebral artery (P2-PCA) was obtained bilaterally as previously published.<sup>10,13</sup>

**BOLD-CVR Data Acquisition.** The BOLD-CVR was performed on a 3T Skyra VD13 (Siemens Healthcare) scanner with a 32-channel head coil. The following imaging parameters were used for the calculation of the whole-brain BOLD volumes: an axial 2D echo-planar imaging BOLD fMRI sequence planned on the

anterior/posterior commissure line plus 20° on a sagittal image with voxel size 3 × 3 × 3 mm<sup>3</sup>, acquisition of matrix 64 × 64 × 35 slices with ascending interleaved acquisition, slice gap 0.3 mm, GRAPPA factor 2 with 32 ref. lines, TR of 2000 ms, TE 30 ms, flip angle 85°, bandwidth 2368 Hz/Px, and a field of view 192 × 192 mm. For every subject, 200 volumes were acquired according to the previously published detailed acquisition protocol.<sup>9</sup>

**BOLD-CVR Data Preprocessing.** Image analysis was conducted by using the software Statistical Parametric Mapping (SPM 12). With the CVR definition as a percentage BOLD signal change per mm Hg CO<sub>2</sub>, CVR calculation was done by using a well-established method. BOLD-CVR was computed by determining the ratio of BOLD signal alteration per mm Hg of CO<sub>2</sub> by analyzing the gradient of a linear least-squares adjustment between the time courses of BOLD signal and CO<sub>2</sub> variations during the BOLD scan as previously published.<sup>9</sup>

**NOVA Data Acquisition.** The NOVA-qMRA data were acquired according to a protocol by using 2D phase-contrast MR sequences, as described in a previous study.<sup>14</sup> The data were acquired by using 3T Skyra VD13 (Siemens Healthcare) and a separate workstation with the NOVA software for the planning/calculation of MR flow measurements. For a simplified quantitative measurement of the hemispheric perfusion and collateral pathways, we only took into consideration the following 6 vessels: A2-ACA, M1-MCA, and P2-PCA bilaterally. The hemispheric flow was calculated as VFR of A2-ACA + M1-MCA + P2-PCA.<sup>10</sup>

#### **BOLD-CVR and qMRA-NOVA Correlation Analysis**

For this study, the following correlations were analyzed:

- P2-PCA VFR versus MCA BOLD-CVR
- P2-PCA VFR versus hemispheric BOLD-CVR

The hemispheric BOLD-CVR refers to the hemispheric BOLD-CVR of the affected hemisphere, including the BOLD-CVR values of the ACA, MCA, and PCA vascular territories.<sup>2</sup>

**Statistical Analysis.** R Studio Version 4.4.0 with the packages readxl, ggplot2, lme4, ggeffects, and lmerTest was used for the statistical analysis of the study data. A Shapiro-Wilk Test has been used to confirm the normal distribution. All normally distributed continuous variables are presented as mean ± SD and non-normally distributed variables as median with interquartile range, and categorical values are displayed as frequencies. A linear mixed-effects model was used to analyze the correlation between BOLD-CVR and blood flow NOVA values. In this model, BOLD-CVR was the independent variable, qMRA-NOVA the dependent one, and the side of measurement (ipsilateral/contralateral) was seen as a random effect. The strength of association was defined as low for 0.1 < |r| < 0.3, moderate for 0.3 < |r| < 0.5, and strong for 0.5 < |r| < 1.0. Finally, for the comparison of the original model with the model being adjusted for confounders (age, sex, smoking, and hypertension), the coefficients of these models were compared, and if they did not differ more than 10%, the parameters were not considered as confounders.

**Table 1: Patient characteristics**

Baseline Characteristics	n = 37
Age (mean ± SD), y (range)	64 ± 11 (37–82)
Male, n (%)	28 (76)
Smoking, n (%)	21 (57)
Hypertension, n (%)	21 (57)
Dyslipidemia, n (%)	22 (59)
Diabetes mellitus	9 (24)
Coronary artery disease, n (%)	6 (16)
Peripheral arterial occlusive disease, n (%)	3 (8.1)
Atrial fibrillation, n (%)	6 (16)
Event classification, n (%)	
• Ischemic stroke, n (%)	34 (92)
• Transient ischemic attack, n (%)	2 (5.4)
• Retinal artery occlusion, n (%)	1 (2.7)
Degree of contralateral stenosis, n (%)	
• No contralateral stenosis	19 (51)
• Low (<50%), n (%)	9 (24)
• Moderate (50%–69%), n (%)	8 (22)
• High (≥70%), n (%)	1 (2.7)

**Table 2: Summary BOLD-CVR values**

BOLD-CVR Values	Mean ± SD
(%BOLD signal change/mm Hg CO <sub>2</sub> )	
BOLD-CVR whole brain	0.09 ± 0.09
BOLD-CVR gray matter	0.11 ± 0.09
BOLD-CVR white matter	0.05 ± 0.08
BOLD-CVR ipsilateral hemisphere	0.07 ± 0.10
BOLD-CVR contralateral hemisphere	0.12 ± 0.07
BOLD-CVR ipsilateral MCA territory	0.07 ± 0.10
BOLD-CVR contralateral MCA territory	0.12 ± 0.07

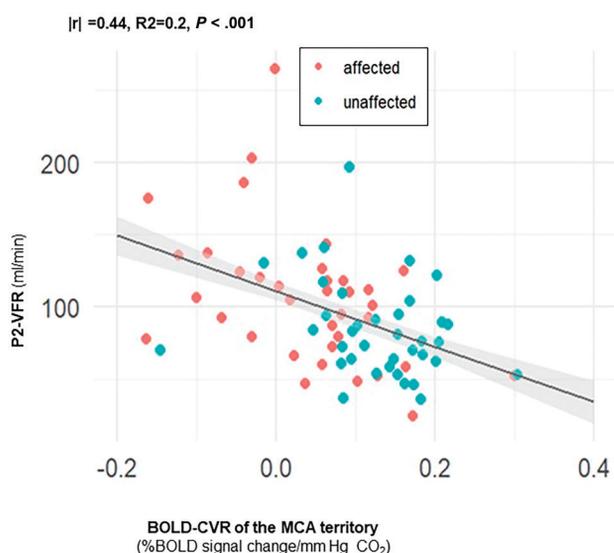
## **RESULTS**

### **Patient Characteristics**

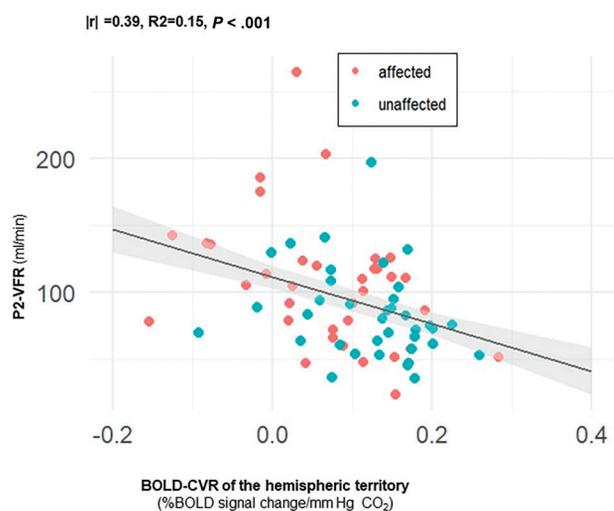
Thirty-seven patients with symptomatic ICA occlusion were included in this study. The mean age was 64 ± 11 years (range of 37–82 years), and 28 were male. The median time between the 2 examinations was 1 day (range 0–12 days). Table 1 shows the patient baseline characteristics. In Table 2, the mean BOLD-CVR values in the different territories are presented, where ipsilateral refers to the affected side (side of the internal carotid artery occlusion). The mean ± SD of P2-VFR for the included cohort is 111 ± 54.0 mL/min on the ipsilateral (affected) side and 81.6 ± 28.6 mL/min on the unaffected side.

### **Correlation Analysis**

A correlation analysis between 1) P2-VFR and BOLD-CVR of the affected MCA territory and 2) P2-VFR and BOLD-CVR of the affected hemisphere was performed. For the comparison between the P2-VFR and BOLD-CVR values of MCA territory, the linear mixed-effects model (|r| = 0.44, R<sup>2</sup> = 0.2, P < .001; Fig 1) shows a moderate correlation. In the correlation analysis for the affected side (Fig 1), it is visible that a BOLD-CVR value of 0 comes along with a P2-VFR of 111 mL/min, and with each BOLD-CVR decrease of 0.1, the P2-VFR increases with 19.2 mL/min. The resulting adjusted R<sup>2</sup> value of 0.20 means that 20% of the variance of the P2-VFR can be explained by the BOLD-CVR of the MCA territory. The r value of −0.44 indicates the inverse (ie, negative) correlation between the P2-VFR and BOLD-CVR values of the MCA territory.



**FIG 1.** Correlation analysis between NOVA-qMRA-derived P2-VFR and BOLD-CVR of the MCA territory. Linear mixed-effects model. A moderate correlation is seen with  $r = 0.44$ ,  $R^2 = 0.2$ , and  $P < .001$ . The adjusted  $R^2$  value of 0.20 means that 20% of the variance of the P2-VFR can be explained by the BOLD-CVR of the MCA territory. The affected side refers to ipsilateral to ICA stenosis.



**FIG 2.** Correlation analysis between NOVA-qMRA-derived P2-VFR and hemispheric BOLD-CVR. At the top, the linear mixed-effects model considering both hemispheres is shown. This shows a moderate correlation with  $r = 0.39$ ,  $R^2 = 0.15$ , and  $P < .001$ . The adjusted  $R^2$  value of 0.15 means that 15% of the variance of the P2-VFR can be explained by the BOLD-CVR of the MCA territory. The affected side refers to ipsilateral to ICA stenosis.

In comparing the P2-VFR with hemispheric BOLD-CVR, the linear mixed-effects model ( $|r| = 0.39$ ,  $R^2 = 0.15$ ,  $P < .001$ ; Fig 2) reveals a moderate correlation. The resulting adjusted  $R^2$  values of 0.15 imply that approximately 15% of the variance in the P2-VFR can be accounted for by the hemispheric BOLD-CVR in these respective models. The  $r$  value of  $-0.39$  indicates the inverse (ie, negative) correlation between P2-VFR and the hemispheric BOLD-CVR.

### Correlation Analysis with Adjustment for Confounders

We adjusted for the following 4 variables: age, sex, smoking, and hypertension, as well as a combination of these confounders. The adjustment was applied to the comparison between P2 flow in NOVA and MCA BOLD-CVR, which showed the strongest correlation. Since the resulting confounder coefficients varied by less than the commonly used 10% cutoff,<sup>15</sup> it can be concluded that the confounders had no significant influence on the results.

In Fig 3, an illustrative case of a patient presenting with symptomatic ICA occlusion is demonstrated. The BOLD-CVR map shows severely impaired BOLD-CVR in the left ACA and MCA territory, including negative CVR-values indicating the steal phenomenon. The NOVA-qMRA report shows that the P2 flow is nearly double the amount on the affected left side than on the unaffected side, indicating activation of leptomeningeal collaterals in this severely impaired hemodynamic situation.

### DISCUSSION

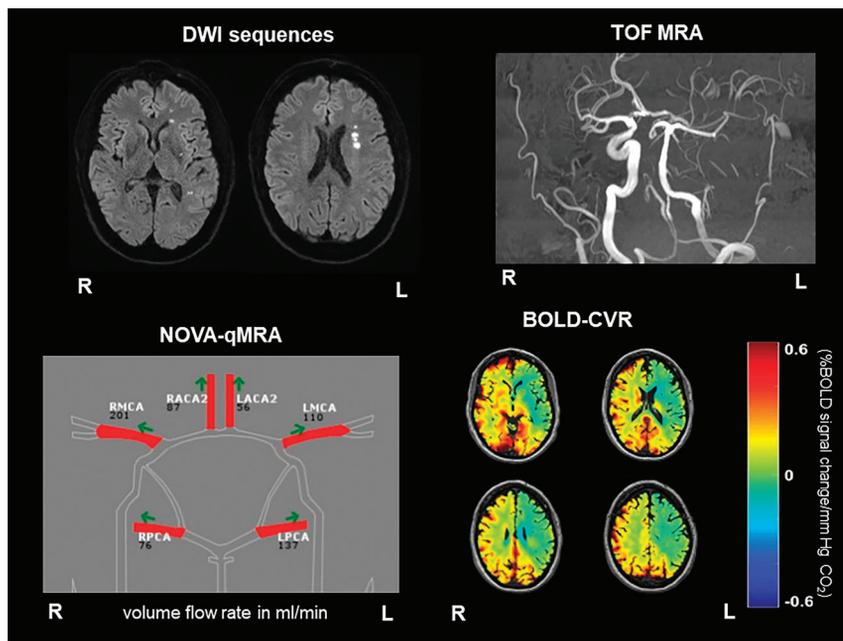
Previous studies using transcranial Doppler (TCD) have shown that increased P2 segment flow velocity in the posterior cerebral artery serves as a surrogate marker for leptomeningeal collateral activation in patients with symptomatic unilateral ICA occlusion. Patients with activated secondary collaterals demonstrated more severe impairment in BOLD-CVR values within the ipsilateral hemisphere. Notably, activation of leptomeningeal collaterals was strongly associated with severely impaired BOLD-CVR in the ipsilateral hemisphere.<sup>8</sup> Additionally, increased ipsilateral PCA-P2 systolic flow velocity, as measured by TCD, was found to correlate independently with hemodynamic failure based on BOLD-CVR measurements.<sup>16</sup> To date, no other mechanisms have been identified that could explain the increased flow observed in these vessels. These findings provided the rationale for focusing on P2-PCA VFR in the present study.

Our data show a moderate correlation between increased P2-VFR and impaired BOLD-CVR in both the ipsilateral hemisphere and the ipsilateral MCA territory. The results of the linear mixed-effects model further indicate that decreased cerebrovascular reserve capacity in the ipsilateral (affected) MCA territory and ipsilateral hemisphere is associated with increased quantitative flow in PCA-P2 vessels, as measured by NOVA-qMRA. This suggests that leptomeningeal collateral activation occurs in severe hemodynamic situations.

This finding aligns with the previous findings having identified ipsilateral leptomeningeal collateral activation as the sole collateral pathway associated with severely impaired BOLD-CVR.<sup>8</sup>

### Relevance of BOLD-CVR and qMRA-NOVA Imaging Techniques

BOLD cerebrovascular reactivity is increasingly used as a hemodynamic imaging technique to assess the cerebrovascular reserve capacity in patients with symptomatic atherosclerotic large artery occlusion.<sup>2,10</sup> In patients with symptomatic ICA occlusion, the cerebrovascular reserve capacity can be altered due to chronic vasodilation and the blood vessel lacking the ability to further reduce its downstream resistance after vasodilatory challenge.<sup>17,18</sup> Furthermore, the volume flow rate can be altered due to the collateral pathways that will be activated in compromised tissue.<sup>4,19</sup>



**FIG 3.** An illustrative case of a patient presenting with symptomatic ICA occlusion. The BOLD-CVR map shows severely impaired BOLD-CVR in the left ACA and MCA territory, including negative CVR-values indicating the steal phenomenon. The NOVA-qMRA report shows that the P2 flow is nearly double the amount on the affected left side than on the unaffected right side.

Since these 2 imaging modalities—NOVA-qMRA-derived VFR and BOLD-CVR—have only been compared in the context of Moyamoya vasculopathy,<sup>20</sup> where they were found effective in assessing hemodynamic and flow impairments before and after cerebral bypass surgery, the findings of this present study open new avenues for the clinical application of these advanced quantitative MRI modalities to assess cerebral hemodynamic and collateral vessel status in patients with symptomatic ICA occlusion.

In clinical practice, combining BOLD-CVR and NOVA-qMRA could aid in stratifying patients by stroke risk. This approach is particularly valuable for those with significantly impaired CVR and elevated P2 flow, who may require closer monitoring and, in selected cases, endovascular or surgical revascularization to reduce the risk of recurrent stroke. According to the literature, these patients with symptomatic ICA occlusion have an overall risk of recurrent ischemic stroke events between 10% and 22%. When considering hemodynamic impairment, this risk increases to 44%, highlighting the importance of proper identification and management of these patients in clinical practice.<sup>21-23</sup>

**Impact of P2-VFR Measurement in Context of Hemodynamic BOLD-CVR Study.** With our study, we show a moderately significant correlation between P2-VFR and BOLD-CVR of the affected MCA territory for the first time. Here, the decreased BOLD-CVR in the affected MCA territory comes along with an increased quantitative flow in P2-PCA vessels, measured as P2-VFR by NOVA-qMRA. Quantitatively, one can say that a BOLD-CVR value of 0 comes along with a P2-VFR of 111 mL/min, and by each BOLD-CVR decrease of 0.1% change in BOLD signal per mm Hg change in CO<sub>2</sub>, the P2-VFR increases by 19.2 mL/min. The findings support what Sebök et al<sup>8</sup> have demonstrated in a previous study, namely a correlation between an increased

ipsilateral Doppler-derived PCA-P2 systolic flow velocity (a surrogate for activation of leptomeningeal collaterals) with BOLD-CVR-based hemodynamic impairment. In this study, TCD-derived leptomeningeal collateral activation was defined as a flow velocity increase of >30% in the ipsilateral P2 segment of the posterior cerebral artery compared with the contralateral P2 segment.<sup>7,24</sup> That leptomeningeal collateral activation for TCD implies the leptomeningeal collateral pathway supplied by the posterior circulation.<sup>25</sup> Here, patients with symptomatic ICA occlusion and activation of only primary collateral pathways (anterior communicating and posterior communicating artery collaterals) did not exhibit an impaired BOLD-CVR. Only activation of leptomeningeal secondary collaterals was associated with an impaired BOLD-CVR and suggested that the activation of leptomeningeal collaterals illustrates a state of maximal hemodynamic compromise

and can, therefore, be taken as an indicator of how severe the disease is.<sup>8</sup> A very recent study further confirmed this finding of leptomeningeal activation in severe carotid stenosis.<sup>26</sup>

Analyzing the parameters of our comparison (P2-flow and BOLD-CVR) in detail, one point worth mentioning is the outlier with a very high P2 flow value. After an in-depth analysis of the medical history, no reason for this high P2 flow was found. Since the parameters for the comparison between P2-PCA flow and MCA BOLD-CVR remained very similar, it can be concluded that P2-VFR derived by NOVA-qMRA has an independent correlation with BOLD-CVR of the affected MCA territory. The exemplary patient in Fig 3 illustrates the above-mentioned pathophysiological mechanism of impaired BOLD-CVR in the affected MCA territory and increased P2 flow, interpreted as a sign of leptomeningeal collateral activation.

Having a look at the correlation between P2-VFR and hemispheric BOLD-CVR, there is still a moderate, significant correlation that is less strong than the above comparison of P2-VFR and BOLD-CVR in the affected MCA territory. The fact that this correlation is less strong could be explained by the MCA territory being the most severely affected by ICA occlusion, as well as the largest vascular territory. Consequently, in many patients, impaired BOLD-CVR can only be found in the MCA territory, while the ACA and PCA territories still have an unaffected or less affected CVR. This results in higher absolute BOLD-CVR values in the affected hemisphere compared with the affected MCA territory.

**Clinical Implication and Future Considerations.** The findings of this study are clinically relevant since many patients do not experience any symptoms but are at high risk of developing a stroke.<sup>27,28</sup> BOLD-CVR and qMRA-NOVA measurements facilitate a detailed assessment of brain hemodynamics and collateral flow status, supporting clinicians in evaluating an individual patient's

risk.<sup>10,13</sup> These assessments help identify patients with exhausted cerebral hemodynamics and increased leptomeningeal collateral effort, making them candidates for endovascular or surgical revascularization strategies aimed at reducing the future risk of recurrent stroke.<sup>10,29</sup> A larger study population would allow dividing the cohort into subgroups to assess the effect of elapsed time between the date of symptoms and the examination. Since leptomeningeal collaterals need some time to be activated in contrast to primary collaterals, which can be rapidly activated, this elapsed time difference could influence the number of leptomeningeal collaterals activated and, therefore, the amount of P2 flow. Furthermore, utilizing these parameters in follow-up imaging may aid in monitoring disease progression and the effectiveness of interventions aimed at maintaining cerebral perfusion. Furthermore, these advanced imaging techniques can also be utilized in future clinical trials to evaluate the effectiveness of endovascular and surgical revascularization strategies in patients with ICA occlusion.

**Limitations.** This study was conducted as a single-center investigation. Despite the small sample size, we were able to obtain significant results for the key correlation between P2 flow and MCA BOLD-CVR. However, a notable limitation of this study is the potential for selection bias, as the BOLD-CVR examinations were specifically performed on patients with internal carotid occlusion who were already suspected of having hemodynamic impairment. Additionally, the time interval between the 2 examinations poses another limitation. Finally, within our study population, only a limited number of patients were identified as having a fetal posterior communicating artery, preventing any analysis that could differentiate between normal and fetal posterior communicating artery configurations. Furthermore, we did not include data on vertebral artery variations, osteal stenosis, or variations in the origin of the vertebral arteries because of the limited data available. Asymptomatic patients were not included in this study, but investigating this cohort would be an interesting question for future research. Additionally, a prospective blinded study with repeated measurements at multiple time points would help validate our findings. Finally, the findings in this study are based on BOLD-CVR with CO<sub>2</sub> challenges; however, additional MR-based perfusion techniques could be used to assess cerebrovascular reserve in future investigations.

## CONCLUSIONS

In patients with recently symptomatic unilateral ICA occlusion, a moderately significant correlation was observed between PCA-P2 volume flow rate and BOLD-CVR in both the ipsilateral MCA territory and the ipsilateral hemisphere. This suggests that impaired BOLD-CVR in these areas may be associated with an increased PCA-P2 volume flow rate, likely due to the activation of secondary leptomeningeal collaterals in severe hemodynamic conditions. These findings support the importance of advanced imaging techniques in detecting early hemodynamic compromise in patients at risk of (recurrent) ischemic events. Such techniques may assist clinicians in implementing timely revascularization interventions and developing personalized treatment strategies.

**Disclosure forms** provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Malhotra K, Goyal N, Tsvigoulis G. **Internal carotid artery occlusion: pathophysiology, diagnosis, and management.** *Curr Atheroscler Rep* 2017;19:41 CrossRef Medline
2. Sebök M, van Niftrik CHB, Winkhofer S, et al. **Mapping cerebrovascular reactivity impairment in patients with symptomatic unilateral carotid artery disease.** *J Am Heart Assoc* 2021;10:e020792 CrossRef Medline
3. Bae YJ, Jung C, Kim JH, et al. **Quantitative magnetic resonance angiography in internal carotid artery occlusion with primary collateral pathway.** *J Stroke* 2015;17:320–26 CrossRef Medline
4. Ruland S, Ahmed A, Thomas K, et al. **Leptomeningeal collateral volume flow assessed by quantitative magnetic resonance angiography in large-vessel cerebrovascular disease.** *J Neuroimaging* 2009;19:27–30 CrossRef Medline
5. Sebök M, van der Wouden F, Mader C, et al. **Hemodynamic failure staging with blood oxygenation level-dependent cerebrovascular reactivity and acetazolamide-challenged ((15)O)-H(2)O-positron emission tomography across individual cerebrovascular territories.** *J Am Heart Assoc* 2023;12:e029491 CrossRef Medline
6. Fierstra J, van Niftrik C, Warnock G, et al. **Staging hemodynamic failure with blood oxygen-level-dependent functional magnetic resonance imaging cerebrovascular reactivity: a comparison versus gold standard ((15)O)-H(2)O-positron emission tomography.** *Stroke* 2018;49:621–29 CrossRef Medline
7. Baumgartner RW, Baumgartner I, Mattle HP, et al. **Transcranial color-coded duplex sonography in unilateral flow-restrictive extracranial carotid artery disease.** *AJNR Am J Neuroradiol* 1996;17:777–83 Medline
8. Sebök M, Niftrik C, Lohaus N, et al. **Leptomeningeal collateral activation indicates severely impaired cerebrovascular reserve capacity in patients with symptomatic unilateral carotid artery occlusion.** *J Cereb Blood Flow Metab* 2021;41:3039–51 CrossRef Medline
9. Sebök M, van Niftrik CHB, Piccirelli M, et al. **BOLD cerebrovascular reactivity as a novel marker for crossed cerebellar diaschisis.** *Neurology* 2018;91:e1328–37 CrossRef Medline
10. Sebök M, Esposito G, Niftrik C, et al. **Flow augmentation STA-MCA bypass evaluation for patients with acute stroke and unilateral large vessel occlusion: a proposal for an urgent bypass flowchart.** *J Neurosurg* 2022;137:1047–55 CrossRef Medline
11. Connolly F, Röhl JE, Lopez-Prieto J, et al. **Pattern of activated pathways and quality of collateral status in patients with symptomatic internal carotid artery occlusion.** *Cerebrovasc Dis* 2019;48:244–50 CrossRef Medline
12. Ferguson GG, Eliasziw M, Barr HW, et al. **The North American Symptomatic Carotid Endarterectomy Trial: surgical results in 1415 patients.** *Stroke* 1999;30:1751–58 CrossRef Medline
13. Sebök M, Höbner LM, Fierstra J, et al. **Flow-augmentation STA-MCA bypass for acute and subacute ischemic stroke due to internal carotid artery occlusion and the role of advanced neuroimaging with hemodynamic and flow-measurement in the decision-making: preliminary data.** *Quant Imaging Med Surg* 2024;14:777–88 CrossRef Medline
14. Zhao M, Charbel FT, Alperin N, et al. **Improved phase-contrast flow quantification by three-dimensional vessel localization.** *Magn Reson Imaging* 2000;18:697–706 CrossRef Medline
15. Budtz-Jørgensen E, Keiding N, Grandjean P, et al. **Confounder selection in environmental epidemiology: assessment of health effects of prenatal mercury exposure.** *Ann Epidemiol* 2007;17:27–35 CrossRef Medline
16. van Niftrik CHB, Sebök M, Wegener S, et al. **Increased ipsilateral posterior cerebral artery P2-segment flow velocity predicts hemodynamic impairment.** *Stroke* 2021;52:1469–72 CrossRef Medline
17. Haight TJ, Bryan RN, Erus G, et al. **Vascular risk factors, cerebrovascular reactivity, and the default-mode brain network.** *Neuroimaging* 2015;115:7–16 CrossRef Medline
18. Yamauchi H, Kudoh T, Sugimoto K, et al. **Pattern of collaterals, type of infarcts, and haemodynamic impairment in carotid artery**

- occlusion.** *J Neurol Neurosurg Psychiatry* 2004;75:1697–701 CrossRef Medline
19. Santhumayor BA, White TG, Golub D, et al. **Impact of cerebral revascularization on pial collateral flow in patients with unilateral moyamoya disease using quantitative magnetic resonance angiography.** *Neurosurgery* 2024;95:596–604 CrossRef Medline
  20. Garbani Nerini L, Bellomo J, Hobner LM, et al. **BOLD cerebrovascular reactivity and NOVA quantitative MR angiography in adult patients with moyamoya vasculopathy undergoing cerebral bypass surgery.** *Brain Sci* 2024;14:20240729 CrossRef Medline
  21. Derdeyn CP. **Hemodynamics and oxygen extraction in chronic large artery steno-occlusive disease: clinical applications for predicting stroke risk.** *J Cereb Blood Flow Metab* 2018;38:1584–97 CrossRef Medline
  22. Baron JC. **Using PET to identify carotid occlusion patients at high risk of subsequent stroke: further insights.** *J Neurol Neurosurg Psychiatry* 2004;75:1659–60 CrossRef Medline
  23. Donahue MJ, Achten E, Cogswell PM, et al. **Consensus statement on current and emerging methods for the diagnosis and evaluation of cerebrovascular disease.** *J Cereb Blood Flow Metab* 2018;38:1391–417 CrossRef Medline
  24. Reinhard M, Müller T, Guschlbauer B, et al. **Dynamic cerebral autoregulation and collateral flow patterns in patients with severe carotid stenosis or occlusion.** *Ultrasound Med Biol* 2003;29:1105–13 CrossRef Medline
  25. Liebeskind DS. **Collateral circulation.** *Stroke* 2003;34:2279–84 CrossRef Medline
  26. Holmgren M, Henze A, Wahlin A, et al. **Phase-contrast magnetic resonance imaging of intracranial and extracranial blood flow in carotid near-occlusion.** *Neuroradiology* 2024;66:589–99 CrossRef Medline
  27. Derdeyn CP, Grubb RL, Jr., Powers WJ. **Cerebral hemodynamic impairment: methods of measurement and association with stroke risk.** *Neurology* 1999;53:251–59 CrossRef Medline
  28. Gupta A, Giambrone AE, Gialdini G, et al. **Silent brain infarction and risk of future stroke: a systematic review and meta-analysis.** *Stroke* 2016;47:719–25 CrossRef Medline
  29. Sebök M, Höbner LM, Grob A, et al. **Flow capacity of a superficial temporal artery as a donor in a consecutive series of 100 patients with superficial temporal artery-middle cerebral artery bypass.** *J Neurosurg* 2025;142:62–69 CrossRef Medline

# Quantitative Shape Irregularity and Density Heterogeneity of Preoperative Hematoma Can Predict Rebleeding following Minimally Invasive Catheter Evacuation for Intracerebral Hemorrhage

 Kaijiang Kang, Zeqiang Ji, Yang Du, Guangshuo Li, Jing Yan, Zeyu Ding, Yiming Shi, Yanfang Liu, Jianwei Wu, and  Xingquan Zhao

## ABSTRACT

**BACKGROUND AND PURPOSE:** Postoperative rebleeding is a critical factor associated with poor outcomes in patients with intracerebral hemorrhage (ICH) who undergo minimally invasive catheter evacuation (MICE) followed by thrombolysis. This study aimed to explore the association between quantitative shape irregularity and density heterogeneity of preoperative hematoma and rebleeding after MICE.

**MATERIALS AND METHODS:** We analyzed patients with ICH who underwent MICE between February 2021 and January 2024. The surface regularity index (SRI) and density coefficient of variation (DCV) of the hematomas were obtained based on preoperative CT by using 3D Slicer software. Postoperative rebleeding was defined as a hematoma increase of >6 mL or >33% compared with the previous CT. The predictive value of shape irregularity (reflected by SRI) and density heterogeneity (reflected by DCV) for postoperative rebleeding were comprehensively analyzed.

**RESULTS:** In total, 240 patients were included, of whom 45 (18.8%) experienced postoperative rebleeding. Patients with postoperative rebleeding exhibited lower SRI (37.2 versus 51.4,  $P = .001$ ) and higher DCV (13.8% versus 11.7%,  $P < .001$ ) after adjusting for preoperative hematoma volume, surface area, standard deviation of hematoma density, intraventricular hemorrhage (IVH), hematoma expansion (HE), time period from onset to surgery, and catheter misplacement. The combination of SRI, DCV, IVH, and HE demonstrated optimal discrimination in predicting postoperative rebleeding, with an area under the curve (AUC) and 95% CI of 0.880 (0.824–0.935).

**CONCLUSIONS:** Hematoma shape irregularity and density heterogeneity are risk factors for rebleeding after MICE for ICH. SRI and DCV can be used to identify individuals at high risk of postoperative rebleeding.

**ABBREVIATIONS:** AUC = area under the curve; DCV = density coefficient of variation; HE = hematoma expansion; ICH = intracerebral hemorrhage; IQR = interquartile range; IVH = intraventricular hemorrhage; MICE = minimally invasive catheter evacuation; OR = odds ratio; ROC = receiver operating characteristic; SRI = surface regularity index

Intracerebral hemorrhage (ICH) is a devastating form of stroke that accounts for 10% to 30% of all strokes worldwide, with a 30-day mortality rate of 30% to 40%; most patients survive with disabilities, leading to a significant burden.<sup>1,2</sup> Hematoma evacuation has long been thought to reduce hematoma volume, decrease intracranial

pressure, and alleviate secondary brain injury by facilitating early removal of the hematoma.<sup>3</sup> However, the International Surgical Trial in Intracerebral Hemorrhage (STICH) and Surgical Trial in Lobar Intracerebral Haemorrhage (STICH-II) demonstrated no benefit in functional outcomes associated with hematoma removal from early craniotomy when compared with initial conservative treatment.<sup>4,5</sup> The Early Minimally-invasive Removal of IntraCerebral Hemorrhage (ENRICH) trial demonstrated that minimally invasive hematoma evacuation based on the BrainPath-Myriad system (NICO) achieved better long-term functional outcomes for patients with lobar hemorrhage, but still not for those with deep hemorrhage.<sup>6</sup>

Minimally invasive catheter evacuation (MICE) for hematoma evacuation has been widely used as an accessible and practical

Received October 13, 2024; accepted after revision December 20.

From the Department of Neurology (K.K., Z.J., Y.D., G.L., J.Y., Z.D., Y.S., Y.L., J.W., X.Z.), Beijing Tiantan Hospital, Capital Medical University, Beijing, China; China National Clinical Research Center for Neurological Diseases (K.K., Z.J., Y.D., G.L., J.Y., Z.D., Y.S., Y.L., J.W., X.Z.), Beijing, China; Center of Stroke (K.K., Z.J., Y.D., G.L., J.Y., Z.D., Y.S., Y.L., J.W., X.Z.), Beijing Institute for Brain Disorders, Beijing, China; Laboratory for Clinical Medicine (X.Z.), Capital Medical University, Beijing, China; and Research Unit of Artificial Intelligence in Cerebrovascular Disease (X.Z.), Chinese Academy of Medical Sciences, Beijing, China.

Kaijiang Kang, Zeqiang Ji, and Yang Du contributed equally to this article.

Prof. Xingquan Zhao, Jianwei Wu and Yanfang Liu are co-corresponding authors.

This work was supported by the National Natural Science Foundation of China (Grant number: 82471489 to K.K. and 82371320 to X.Z.), Health China-BuChang ZhiYuan Public Welfare Projects for Heart and Brain Health (Grant number: HIGHER2023074 to K.K.), and Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (Grant number: 2019-I2M-5-029 to X.Z.).

Please address correspondence to Xingquan Zhao, Beijing Tiantan Hospital, Capital Medical University, Department of Neurology, No.119 South Fourth Ring West Rd, Fengtai District Beijing, Beijing 100070, China; e-mail: zxq@vip.163.com

 Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8680>

## SUMMARY

**PREVIOUS LITERATURE:** MICE for hematoma evacuation has been widely used as an accessible and practical approach for ICH. However, postoperative rebleeding has been indicated to be one of the key factors affecting the benefit of surgery. Prior research has indicated that several NCCT signs reflecting the shape irregularity and density heterogeneity of the hematoma were associated with postoperative rebleeding. However, these NCCT signs were qualitative, and the interpretation was significantly subjective, resulting in a suboptimal predictive ability for postoperative rebleeding. This study was investigated to explore the association of hematoma characteristics quantitatively with postoperative rebleeding.

**KEY FINDINGS:** This study indicated that hematoma shape irregularity and density heterogeneity were risk factors for rebleeding after MICE for ICH and demonstrated that SRI and DCV combined with IVH and HE could predict postoperative rebleeding with optimal discrimination.

**KNOWLEDGE ADVANCEMENT:** Based on the findings of this study, SRI and DCV can be used to identify individuals at high risk of postoperative rebleeding. For patients with elevated SRI and DCV, repeated NCCT assessment for delayed surgery decisions can be considered.

approach because of its advantages of shorter operation times and less disruption of unaffected brain tissue compared with conventional craniotomy, which has attracted considerable attention.<sup>7</sup> However, the Minimally Invasive Surgery with Plus Rt-PA for ICH Evacuation Phase III (MISTIE III) trial failed to identify any definite therapeutic improvement in the functional outcomes of MICE compared with conventional treatment, one of the reasons that was indicated was the higher rate of postoperative rebleeding in the MICE group.<sup>8</sup>

Prior research has indicated that several NCCT signs reflecting the shape irregularity and density heterogeneity of the hematoma, such as an irregular shape,<sup>9</sup> black hole sign,<sup>10</sup> and blend sign,<sup>11</sup> are all associated with postoperative rebleeding. However, these NCCT signs were qualitative, and the interpretation was significantly subjective, resulting in a suboptimal predictive ability for postoperative rebleeding.<sup>9,10</sup> The 3D Slicer software ([www.slicer.org](http://www.slicer.org)) has previously been utilized in ICH research for hematoma modeling and puncture path selection for MICE and can also provide objective and quantitative measurements of the shape irregularity and density variability of the hematoma.<sup>12,13</sup> Herein, we investigated the association between shape regularity and density variability of preoperative hematomas by using 3D Slicer software with postoperative rebleeding in patients with ICH who underwent MICE.

## MATERIALS AND METHODS

### **Study Population, Clinical Characteristics, Preoperative Imaging Program, NCCT Sign Assessment, MICE Procedure, and Periprocedural Management**

The details of the study population, clinical characteristics, preoperative imaging program, NCCT sign assessment, MICE procedure, and periprocedural management are presented in Supplemental Data.

### **Definition and Measurement of Hematoma Shape and Density**

The hematoma was segmented and reconstructed from the DICOM data of the preoperative CT by using the 3D Slicer software (Version 4.10.1, [www.slicer.org](http://www.slicer.org)) (Fig 1). Hematomas were

semiautomatically identified pixel-by-pixel in each slice, with thresholds ranging from 40 to 100 HU.<sup>14</sup> Finally, the model module was utilized to reconstruct the 3D data by adding all pixels from each slice, and the values of hematoma volume and surface area were directly obtained from the 3D Slicer without smoothing processing. The surface regularity index (SRI) was calculated from hematoma volume and surface area, with values between 0 (fractal hematoma with very irregular surfaces) and 1 (spherical hematoma), by using the following formula<sup>15</sup>:

$$SRI = 6\sqrt{\pi} \left( \frac{\text{hematoma volume}}{\sqrt{(\text{hematoma surface area})^3}} \right) * 100$$

After segmenting and modeling the hematoma, we applied the “statistics” modules to automatically calculate the mean, median, and SD of the CT values of each pixel. The density coefficient of variation (DCV) was calculated to represent the hematoma density heterogeneity.

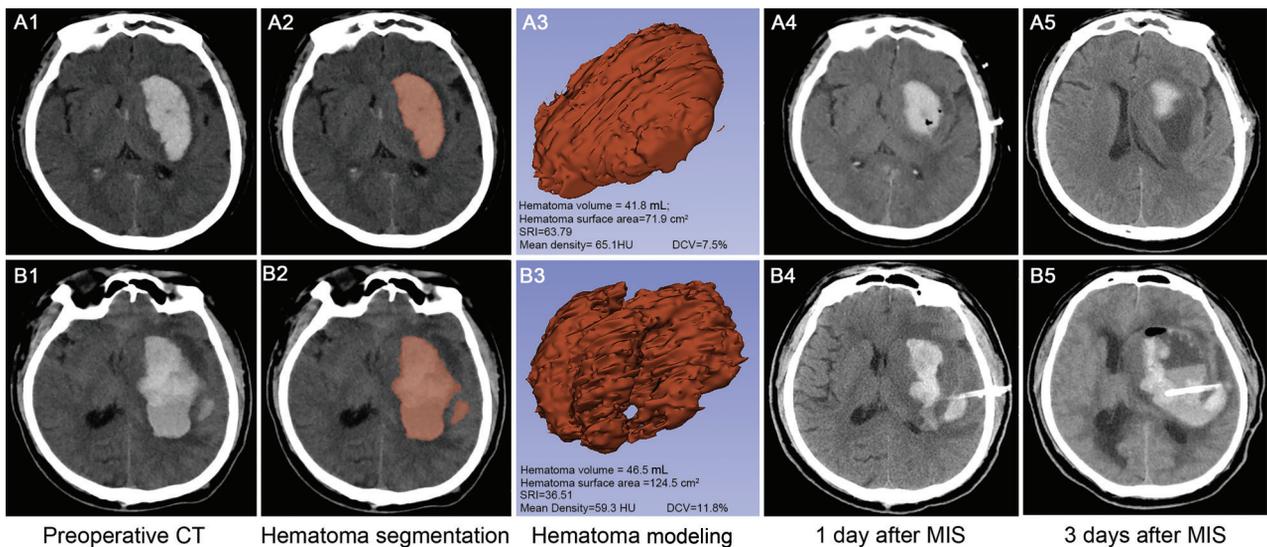
$$DCV = \left( \frac{\text{Standard Deviation}}{\text{Mean}} \right) * 100\%$$

### **Neuroradiological Image Interpretation**

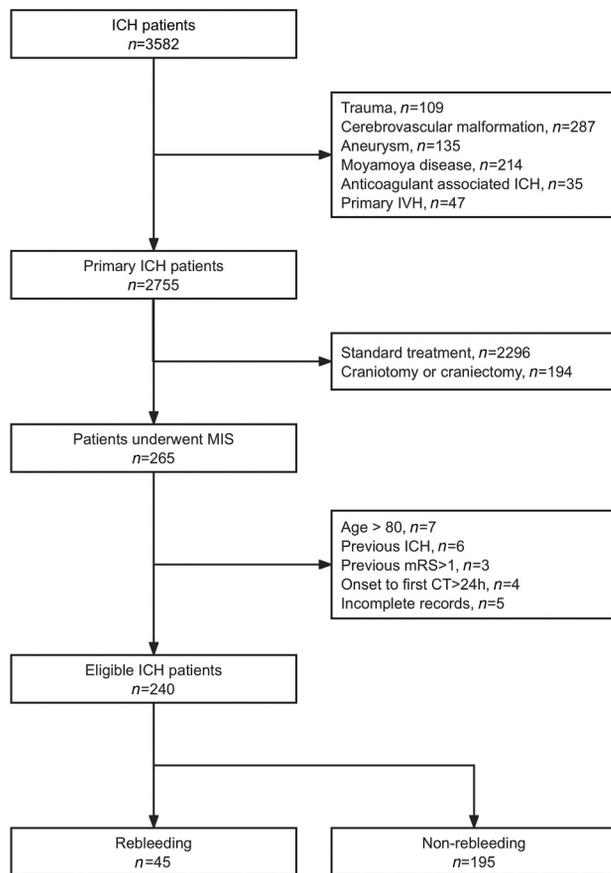
All neuroradiological images were interpreted independently by 2 experienced neuroradiologists (blinded to the outcome), and an interpretation training program was performed for both to ensure consistency of the interpreting standard. After the training program, a consistency test was carried out, and the formal interpretation started when the intraclass correlation coefficient reached above 0.8. The time needed to process each NCCT imaging differed from 5 to 10 minutes. The differences between assessments were resolved by a third senior neuroradiologist (categorical data), or the average of their data (quantitative data) was calculated to reduce the impact of subjective factors.

### **Statistical Analysis**

Continuous variables were expressed as medians (interquartile range [IQR]) or means  $\pm$  SD, which were evaluated by using the Shapiro-Wilk test and compared by using the Mann-Whitney or



**FIG 1.** Hematoma segmentation, modeling, and postoperative outcomes in 2 exemplar cases. A1–A5, Images taken from a patient with ICH who underwent MICE with a high SRI and low DCV of preoperative hematoma, achieving satisfactory hematoma drainage without postoperative rebleeding. B1–B5, Images taken from an patient with ICH who underwent MICE with low SRI and high DCV of preoperative hematoma and experienced rebleeding 3 days following the MICE. A1–A3 and B1–B3, Hematoma segmentation and modeling performed by using the 3D Slicer software based on preoperative CT. The results of the modeled hematoma measurements and calculations (hematoma volume, surface area, SRI, mean density, and DCV) are presented in A3 and B3.



**FIG 2.** Flowchart of patient selection.

*t* test, respectively. The chi-square test was applied to compare categorical variables, which are expressed as numbers (proportions). Variables with  $P < .05$  from the comparison of baseline

characteristics and variables significantly influencing the outcome in previous studies were included in the multivariate logistic regression. The ORs and 95% CI for postoperative rebleeding were calculated subsequently. The discriminative abilities of the selected variables were evaluated by using receiver operating characteristic (ROC) curves with the De Long test for comparison of areas under the curve (AUC). Differences with  $P < .05$  were considered statistically significant for 2-tailed tests. Statistical analyses were performed using a commercial statistical software package (SPSS for Windows, Version 25.0, IBM-SPSS), and the Delong test was conducted using SAS software (Version 9.4; SAS Institute).

## RESULTS

### Patient Baseline Characteristics

From February 2021 to January 2024, a total of 240 patients ( $56.0 \pm 13.9$  years old) were recruited, including 185 (77.1%) men and 55 (22.9%) women. A flowchart of the patient selection process is shown in Fig 2. The median baseline GCS and NIHSS scores were 11 (8–13) and 18 (13–27), respectively. The median periods from ICH onset to baseline CT and MICE surgery were 7.0 (4.0, 13.0) and 48.1 (36.2, 64.9) hours, respectively. The median baseline hematoma volume and ultraearly hematoma growth (uHG) were 36.7 mL (26.9, 53.7) and 6.0 mL/h (2.9, 10.4), respectively. Deep hematomas (including the basal ganglia and thalamus) accounted for 81.2% of all hemorrhages, whereas lobar hematomas accounted for 18.8%. All the baseline characteristics are presented in the Supplemental Data.

### Periprocedural Characteristics

The median preoperative hematoma volume in this study cohort was 41.8 mL (37.2, 64.9), with 77 patients found to be complicated with intraventricular hemorrhage (IVH). The median surface area of preoperative hematoma was 99.0 cm<sup>2</sup> (74.5, 137.5). The average

### Results of the multivariate logistic regression of postoperative rebleeding prediction<sup>a</sup>

	P Value	OR	95% CI
SRI	.001	0.938	0.905–0.973
DCV	<.001	1.907	1.473–2.470
IVH	.028	2.630	1.109–6.235
HE	.025	2.635	1.127–6.162
Onset to surgery	.173	1.011	0.966–1.025
PH volume	.539	1.005	0.989–1.022
PH surface area	.491	0.992	0.971–1.014
PH density SD	.845	1.120	0.359–3.495
Catheter misplacement	.871	1.084	0.410–2.863

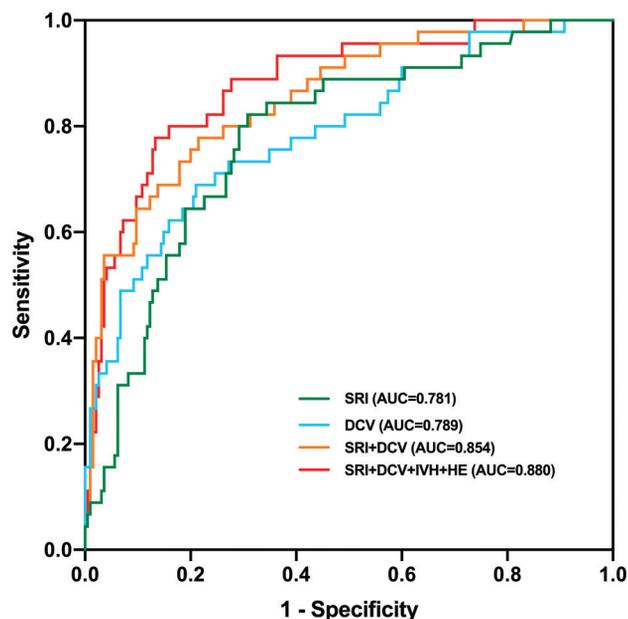
<sup>a</sup>The preoperative hematoma SRI, DCV, IVH, and HE remained significant after adjusting for significant variables in the univariate analysis (including PH volume, surface area, and SD of hematoma density), as well as those influencing the outcomes in previous studies (time from onset to surgery and catheter misplacement). PH = preoperative hematoma.

density of the preoperative hematoma was  $60.4 \pm 3.5$  HU. The mean SRI and DCV of the preoperative hematoma were  $48.7 \pm 15.5$  and  $12.1 \pm 2.0\%$ , respectively. The median depth of the catheter placement was 6.5 cm (6.0, 6.5), and 20.0% (48/240) of catheters were misplaced for the first time. The median volume of the first aspirated hematoma was 19.0 mL, and the drainage was 3.1 days (3.0, 4.0). Of the 240 investigated patients, 45 (18.8%) experienced postoperative rebleeding. The median period from MICE surgery to rebleeding was 3.0 days (3.0, 3.9), ranging from 1.6 to 7 days.

### Factors Associated with Postoperative Rebleeding

Patients with postoperative rebleeding had a larger preoperative hematoma volume (60.0 versus 40.5 mL,  $P < .001$ ) and hematoma surface area (141.2 versus 91.6 cm<sup>2</sup>,  $P < .001$ ), with a higher prevalence of IVH (48.9% versus 28.2%,  $P = .007$ ) and hematoma expansion (HE) (48.9% versus 23.6%,  $P = .001$ ) (Fig 1). The postoperative rebleeding was associated with a lower SRI (37.2 versus 51.4,  $P < .001$ ) and higher DCV (13.8% versus 11.7%,  $P < .001$ ) (Fig 1). The cutoff points of SRI and DCV for predicting postoperative rebleeding were 43.48 and 12.97%, respectively. There were no significant differences in baseline hematoma volume, period from onset to surgery, catheter depth, aspiration volume, or drainage time between patients with and without postoperative rebleeding. Of the 45 patients with postoperative rebleeding, 27 (61.4%) experienced neurologic deterioration, defined as a  $\geq 2$  reduction in GCS or a  $\geq 4$  increase in NIHSS score for non-sedatives/sleeping medications compared with the pre-rebleeding status.<sup>16</sup> We further observed a higher prevalence of poor drainage (91.1% versus 36.4%,  $P < .001$ ) and longer hospital stay (23 versus 15 days,  $P < .001$ ) in patients with postoperative rebleeding than in those without rebleeding (Table).

Multivariate logistic regression analysis indicated that preoperative hematoma SRI (odds ratio [OR] [95% CI]: 0.938 [0.905–0.973],  $P = .001$ ), DCV (OR [95% CI]: 1.907 [1.473–2.470],  $P < .001$ ), IVH (OR [95% CI]: 2.630 [1.109–6.235],  $P = .028$ ), and HE (OR [95% CI]: 2.635 [1.127–6.162],  $P = .025$ ) were independent predictors of postoperative rebleeding, after adjusting for associated variables in univariate analysis (including preoperative hematoma volume, surface area, standard deviation



**FIG 3.** ROC curve analysis of SRI, DCV, and the 2 models for predicting postoperative rebleeding. SRI (AUC [95% CI]: 0.781 [0.708–0.854]) and DCV (AUC [95% CI]: 0.789 [0.710–0.867]) exhibited acceptable predictive abilities for postoperative rebleeding. The combination of SRI and DCV provided significant improvements in rebleeding prediction (AUC [95% CI]: 0.854 [0.791–0.917]) compared with SRI ( $P = .012$ ) and DCV ( $P = .013$ ) alone. In addition, the combination of the 4 selected predictors (SRI, DCV, IVH, and HE) demonstrated optimal discrimination in the prediction of postoperative rebleeding (AUC [95% CI]: 0.880 [0.824–0.935]), showing a significant improvement compared with SRI ( $P = .002$ ) and DCV ( $P = .002$ ).

of hematoma density) and factors that influenced the outcome in a previous study (time from onset to surgery<sup>17</sup> and presence of misplaced catheter<sup>18</sup>) (Table).

### Predictive Analysis

In univariate prediction analysis, SRI (AUC [95% CI]: 0.781 [0.708–0.854]) and DCV (AUC [95% CI]: 0.789 [0.710–0.867]) exhibited acceptable predictive abilities for postoperative rebleeding. The combination of SRI and DCV also provided significant improvements in rebleeding prediction (AUC [95% CI]: 0.854 [0.791–0.917]) compared with SRI ( $P = .012$ ) and DCV ( $P = .013$ ). In addition, the combination of SRI, DCV, IVH, and HE demonstrated optimal discrimination in the prediction of postoperative rebleeding (AUC [95% CI]: 0.880 [0.824–0.935]), showing significant improvements compared with SRI ( $P = .002$ ) and DCV ( $P = .002$ ) (Fig 3).

Additionally, we evaluated the predictive ability of previously reported radiologic signs of postoperative rebleeding in this cohort. NCCT signs (including blend sign, heterogeneous density, hypodensities, black hole sign, island sign, satellite sign, and irregular shape) were observed in all patients. CTA was available for 224 patients (93.3%), with spot signs identified in 32 (14.3%). The median period from onset to CTA was 11 hours (5.0, 20.0). The rebleeding groups had a higher prevalence of heterogeneous density (42.2% versus 27.2%,  $P = .047$ ). However, there was no difference in other NCCT and CTA spot signs between patients

with and without postoperative rebleeding. The AUCs of all reported radiologic signs ranged from 0.502 to 0.588 (Supplemental Data).

## DISCUSSION

To the best of our knowledge, this is the first study to quantitatively and objectively explore the association between shape irregularity and density heterogeneity of hematomas and post-MICE rebleeding. This study demonstrated that SRI and DCV combined with IVH and HE could predict postoperative rebleeding with optimal discrimination.

Minimally invasive hematoma removal has previously been proposed as a strategy to alleviate the mass effect and secondary injuries resulting from the hematoma, reducing the mortality and disability rates of patients with ICH.<sup>3,19</sup> While several meta-analyses have indicated that MICE is associated with better functional outcomes and fewer complications than conservative treatment or craniotomy,<sup>20,21</sup> the MISTIE III trial failed to demonstrate that MICE could improve long-term functional outcomes in patients with ICH compared with conservative treatment.<sup>8</sup> Subsequent research has revealed that poor outcomes after MICE are associated with postoperative rebleeding, along with a high neurologic deterioration rate and large residual hematoma volume resulting from rebleeding.<sup>22,23</sup> In this cohort, the postoperative rebleeding rate was 18.8%, which is comparable with that observed in previous studies (about 20%).<sup>7,8,21</sup> In addition, patients with postoperative rebleeding had a higher rate of poor drainage than those without rebleeding (91.1% versus 36.4%). As such, it is imperative to identify patients at high risk of postoperative rebleeding when selecting patients with ICH for MICE.

Several studies have previously suggested that NCCT signs, such as an irregular hematoma shape, blend sign, and satellite sign, can predict postoperative rebleeding.<sup>9-11</sup> These NCCT signs were initially investigated for HE prediction and can be divided into 2 types according to the essential characteristics of the hematoma<sup>24</sup>: shape irregularity (such as irregular shape,<sup>25</sup> satellite sign,<sup>26</sup> and island sign<sup>27</sup>) and density heterogeneity (such as blend sign,<sup>28</sup> hypodensities,<sup>29</sup> and black hole sign<sup>30</sup>). However, NCCT signs reflecting shape irregularity and density heterogeneity are qualitative and require subjective interpretation. The predictive ability of NCCT signs for postoperative rebleeding has been reported to be suboptimal (AUC of 0.688 for black hole sign,<sup>10</sup> 0.772 for blend sign, and 0.629 for irregular shape<sup>31</sup>). We further evaluated the predictive ability of NCCT and spot signs for rebleeding. However, neither of these signs exhibited acceptable predictive abilities for postoperative rebleeding compared with DCV and SRI (Supplemental Data).

In this study, hematomas were evaluated objectively and quantitatively by using 3D Slicer software based on the pixels of the hematoma, without making any assumptions.<sup>14,32</sup> The SRI was used to describe the irregularity of the hematoma, which was adjusted for the influence of the hematoma volume on the hematoma surface area. The DCV of each pixel was then automatically calculated to represent the hematoma density heterogeneity. This study is the first to use the SRI and DCV to predict post-MICE rebleeding. Additionally, we adopted CT scans with a thickness of 1 mm for segmentation, which is believed to be more accurate

and realistic than the regular approach by using 5 mm CT for hematoma modeling in most studies.

Although the exact underlying pathophysiological mechanism remains unclear, the Fisher domino model<sup>33</sup> is a widely acknowledged theory that explains the association between shape irregularity and HE or rebleeding. According to the Fisher model, HE or rebleeding is caused by the shearing stress of blood vessels around the site of the primary hemorrhage, which contributes to secondary hemorrhage. Irregularly shaped hematomas, according to different definitions, have further been shown to predict HE with acceptable performance.<sup>24,25,34</sup> Nonetheless, the performance of irregular hematomas in predicting postoperative rebleeding has been unsatisfactory. In one previous study, irregular hematoma shape, defined as a categorical scale by Barras et al<sup>25</sup> exhibited suboptimal discriminative ability (AUC = 0.629) for postoperative rebleeding.<sup>31</sup> In the present study, the shape irregularity of the hematoma was reflected by the SRI, which was calculated mathematically with quantitative and objective measurements of hematoma volume and surface area and exhibited acceptable predictive abilities for postoperative rebleeding with an AUC of 0.781.

The DCV of the hematoma, which reflects the density heterogeneity, was also identified as an independent predictor of postoperative rebleeding for the first time in this study, yielding an AUC of 0.789. From our perspective, there are 2 possible explanations for a hematoma with heterogeneous density: first, the hematoma comprises different stages or components of hemorrhage, reflecting the instability of the hematoma; and second, the hemorrhage itself is more dispersed and mixed with the brain tissue. The first case indicated bleeding at different time points, reflecting the instability of the hematoma with a tendency for hematoma expansion. In the second case, the hematoma was mixed with the brain tissue, possibly indicating that the drainage catheter may have penetrated more brain tissue with cerebral vessels during the surgery, resulting in an increased risk of postoperative rebleeding.

Prior studies on the prediction of postoperative rebleeding by using the black hole sign<sup>10</sup> and blend sign<sup>11</sup> were based primarily on baseline (within hours of ICH onset) rather than on preoperative CT. However, the shape and density of hematomas may change over time following ICH onset. In the present study, preoperative CT was applied to evaluate the hematoma that was closest to the hematoma during surgery. The median period from ICH onset to preoperative CT was approximately 2 days, indicating that the time window of shape irregularity and density heterogeneity for predicting postoperative rebleeding could be extended from 6 to 48 hours in clinical practice.

Our results further suggest that IVH and HE were independently associated with postoperative rebleeding. The combination of SRI and DCV with IVH and HE demonstrated optimal discrimination in predicting postoperative rebleeding with an AUC of 0.880.

Notably, the latest ENRICH trial demonstrated that minimally invasive hematoma evacuation based on the BrainPath-Myriad system achieved better functional outcomes at 180 days for patients with lobar hemorrhage.<sup>6</sup> The results of the Minimally Invasive Endoscopic Surgical Treatment With Apollo/Artemis in Patients With Brain Hemorrhage (INVEST) and Artemis in the

Removal of Intracerebral Hemorrhage (MIND) trials (by using the Apollo MIES and Artemis systems [Penumbra], respectively) are also forthcoming. Additionally, several ICH trials on minimally invasive hematoma evacuation are ongoing, including Early Minimally Invasive Image Guided Endoscopic Evacuation of Intracerebral Haemorrhage (EMINENT-ICH), Dutch Intracerebral Hemorrhage Surgery Trial (DIST), and Ultra-Early, Minimally Invasive IntraCerebral Haemorrhage evacuation Versus Standard treatment (EVACUATE). However, the minimally invasive visual devices used in these trials are expensive and less accessible, particularly in developing countries. Nonvisual, minimally invasive hematoma excavation will continue to be an essential surgical procedure in many places, and preoperative hematoma evaluation and patient selection are critical for reducing the risk of postoperative rebleeding.

This study has some limitations that should be considered when interpreting the results. First, the sample size was small, and all the patients were enrolled from a single center; therefore, a potential selection bias may be inevitable. Second, SRI and DCV must be acquired via manual image processing, which may limit their clinical application. Automatic analysis software that calculates and processes these parameters will be developed in future studies. Third, the thrombolytic agent used in this study was urokinase, not alteplase or tenecteplase, as used in other clinical trials. However, urokinase is relatively more commonly used in China, especially in areas that are economically less developed.

## CONCLUSIONS

Hematoma shape irregularity and density heterogeneity are risk factors for rebleeding after MICE for ICH. SRI and DCV can be used to identify individuals at high risk of postoperative rebleeding.

## ACKNOWLEDGMENTS

We thank all of the patients and health care providers who participated in this study.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

- Sheth KN. Spontaneous intracerebral hemorrhage. *N Engl J Med* 2022;387:1589–96 CrossRef Medline
- Tsao CW, Aday AW, Almarzoq ZI, et al. Heart disease and stroke statistics—2022 update: a report from the American Heart Association. *Circulation* 2022;145:e153–e639 CrossRef Medline
- Sondag L, Schreuder F, Boogaarts HD; Dutch ICH Surgery Trial Study Group, part of the CONTRAST Consortium, et al. Neurosurgical intervention for supratentorial intracerebral hemorrhage. *Ann Neurol* 2020;88:239–50 CrossRef Medline
- Mendelow AD, Gregson BA, Fernandes HM, et al. Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the International Surgical Trial in Intracerebral Haemorrhage (STICH): a randomised trial. *Lancet* 2005;365:387–97 CrossRef
- Mendelow AD, Gregson BA, Rowan EN; STICH II Investigators, et al. Early surgery versus initial conservative treatment in patients with spontaneous supratentorial lobar intracerebral haematomas (STICH II): a randomised trial. *Lancet* 2013;382:397–408 CrossRef Medline
- Pradilla G, Ratcliff JJ, Hall AJ; ENRICH Trial Investigators, et al. Trial of early minimally invasive removal of intracerebral hemorrhage. *N Engl J Med* 2024;390:1277–89 CrossRef Medline
- Scaggiante J, Zhang X, Mocco J, et al. Minimally invasive surgery for intracerebral hemorrhage. *Stroke* 2018;49:2612–20 CrossRef Medline
- Hanley DF, Thompson RE, Rosenblum M; MISTIE III Investigators, et al. Efficacy and safety of minimally invasive surgery with thrombolysis in intracerebral haemorrhage evacuation (MISTIE III): a randomised, controlled, open-label, blinded endpoint phase 3 trial. *Lancet* 2019;393:1021–32 CrossRef Medline
- Wang L, Zhang L, Mao Y, et al. Regular-shaped hematomas predict a favorable outcome in patients with hypertensive intracerebral hemorrhage following stereotactic minimally invasive surgery. *Neurocrit Care* 2021;34:259–70 CrossRef Medline
- Shen Z, Wang L, Wu G, et al. Computed tomographic black hole sign predicts postoperative rehemorrhage in patients with spontaneous intracranial hemorrhage following stereotactic minimally invasive surgery. *World Neurosurg* 2018;120:e153–60 CrossRef Medline
- Wu G, Shen Z, Wang L, et al. Post-operative re-bleeding in patients with hypertensive ICH is closely associated with the CT blend sign. *BMC Neurol* 2017;17:131 CrossRef Medline
- Xu X, Chen X, Zhang J, et al. Comparison of the Tada formula with software slicer: precise and low-cost method for volume assessment of intracerebral hematoma. *Stroke* 2014;45:3433–35.20141014. CrossRef Medline
- Peng Y, Xie Z, Chen S, et al. Application effect of head-mounted mixed reality device combined with 3D printing model in neurosurgery ventricular and hematoma puncture training. *BMC Med Educ* 2023;23:670. 20230918. CrossRef Medline
- Zhao B, Jia WB, Zhang LY, et al. 1/2SH: a simple, accurate, and reliable method of calculating the hematoma volume of spontaneous intracerebral hemorrhage. *Stroke* 2020;51:193–201 CrossRef Medline
- Pérez-Beteta J, Molina-García D, Ortiz-Alhambra JA, et al. Tumor surface regularity at MR imaging predicts survival and response to surgery in patients with glioblastoma. *Radiology* 2018;288:218–25 CrossRef Medline
- Dávalos A, Toni D, Iweins F, et al. Neurological deterioration in acute ischemic stroke: potential predictors and associated factors in the European Cooperative Acute Stroke Study (ECASS) I. *Stroke* 1999;30:2631–36 CrossRef Medline
- Kellner CP, Song R, Ali M, et al. Time to evacuation and functional outcome after minimally invasive endoscopic intracerebral hemorrhage evacuation. *Stroke* 2021;52:e536–39 CrossRef Medline
- Awad IA, Polster SP, Carrión-Penagos J; MISTIE III Trial Investigators, et al. Surgical performance determines functional outcome benefit in the Minimally Invasive Surgery Plus Recombinant Tissue Plasminogen Activator for Intracerebral Hemorrhage Evacuation (MISTIE) Procedure. *Neurosurgery* 2019;84:1157–68 CrossRef Medline
- Wang WZ, Jiang B, Liu HM, et al. Minimally invasive craniopuncture therapy vs. conservative treatment for spontaneous intracerebral hemorrhage: results from a randomized clinical trial in China. *Int J Stroke* 2009;4:11–16 CrossRef Medline
- Zhou X, Chen J, Li Q, et al. Minimally invasive surgery for spontaneous supratentorial intracerebral hemorrhage: a meta-analysis of randomized controlled trials. *Stroke* 2012;43:2923–30 CrossRef Medline
- Tang Y, Yin F, Fu D, et al. Efficacy and safety of minimal invasive surgery treatment in hypertensive intracerebral hemorrhage: a systematic review and meta-analysis. *BMC Neurol* 2018;18:136 CrossRef Medline
- Kobata H, Ikeda N. Recent updates in neurosurgical interventions for spontaneous intracerebral hemorrhage: minimally invasive surgery to improve surgical performance. *Front Neurol* 2021;12:703189 CrossRef Medline

23. de Havenon A, Joyce E, Yaghi S, et al. **End-of-treatment intracerebral and ventricular hemorrhage volume predicts outcome: a secondary analysis of MISTIE III.** *Stroke* 2020;51:652–54 CrossRef Medline
24. Morotti A, Arba F, Boulouis G, et al. **Noncontrast CT markers of intracerebral hemorrhage expansion and poor outcome: a meta-analysis.** *Neurology* 2020;95:632–43 CrossRef Medline
25. Barras CD, Tress BM, Christensen S; Recombinant Activated Factor VII Intracerebral Hemorrhage Trial Investigators, et al. **Density and shape as CT predictors of intracerebral hemorrhage growth.** *Stroke* 2009;40:1325–31 CrossRef Medline
26. Shimoda Y, Ohtomo S, Arai H, et al. **Satellite sign: a poor outcome predictor in intracerebral hemorrhage.** *Cerebrovasc Dis* 2017;44:105–12 CrossRef Medline
27. Li Q, Liu Q-J, Yang W-S, et al. **Island sign: an imaging predictor for early hematoma expansion and poor outcome in patients with intracerebral hemorrhage.** *Stroke* 2017;48:3019–25 CrossRef Medline
28. Li Q, Zhang G, Huang YJ, et al. **Blend sign on computed tomography: novel and reliable predictor for early hematoma growth in patients with intracerebral hemorrhage.** *Stroke* 2015;46:2119–23 CrossRef Medline
29. Boulouis G, Morotti A, Brouwers HB, et al. **Association between hypodensities detected by computed tomography and hematoma expansion in patients with intracerebral hemorrhage.** *JAMA Neurol* 2016;73:961–68 CrossRef Medline
30. Li Q, Zhang G, Xiong X, et al. **Black hole sign: novel imaging marker that predicts hematoma growth in patients with intracerebral hemorrhage.** *Stroke* 2016;47:1777–81 CrossRef Medline
31. Wang L, Luo S, Ren S, et al. **Irregular-shaped hematoma predicts postoperative rehemorrhage after stereotactic minimally invasive surgery for intracerebral hemorrhage.** *Front Neurol* 2022;13:727702 CrossRef Medline
32. Kuramatsu JB, Huttner HB, Schwab S. **Advances in the management of intracerebral hemorrhage.** *J Neural Transm (Vienna)* 2013;120 Suppl 1:S35–41 CrossRef Medline
33. Fisher CM. **Pathological observations in hypertensive cerebral hemorrhage.** *J Neuropathol Exp Neurol* 1971;30:536–50 CrossRef Medline
34. Blacquiere D, Demchuk AM, Al-Hazzaa M; PREDICT/Sunnybrook ICH CTA Study Group, et al. **Intracerebral hematoma morphologic appearance on noncontrast computed tomography predicts significant hematoma expansion.** *Stroke* 2015;46:3111–16 CrossRef Medline

# Circle of Willis Variants and Their Association with Outcome after Successful Revascularization of Anterior Large Vessel Occlusion

Aakanksha Sriwastwa,  Michael K. Oswald,  Achala S. Vagal, Stacie L. Demel,  Bin Zhang, Sriharsha Voleti, Arafat Ali, Daniel Morgan, Trevor Thompson, Johnathan Vidovich, Yasmin N. Aziz, and Lily Li-Li Wang



## ABSTRACT

**BACKGROUND AND PURPOSE:** Leptomeningeal collaterals have been associated with better outcomes in large-vessel stroke, but little is known about how the circle of Willis (CoW) collaterals affect stroke outcomes. We aimed to determine the relationship between 3 anatomically distinct CoW subtypes and 90-day outcomes in patients with acute ischemic stroke after successful revascularization via endovascular thrombectomy (EVT).

**MATERIALS AND METHODS:** We performed a retrospective analysis of patients treated with successful EVT for large-vessel occlusion at a comprehensive stroke center between May 2016 and November 2023. The CoW anatomy was trichotomized by using baseline CT angiography as follows: 1) complete circle of Willis (C-CoW), 2) nonisolating incomplete circle of Willis (NI-CoW), and 3) isolating incomplete circle of Willis (I-CoW).  $\chi^2$  and logistic regression analyses were utilized to determine the association of the CoW subtype with 2 coprimary outcomes: the 90-day mRS and 90-day mortality.

**RESULTS:** A total of 465 patients were included in the analysis. Multivariable logistic regression analysis demonstrated a significant association between I-CoW and 90-day mRS compared with NI-CoW (OR [95% CI], 1.83 [1.08–3.09];  $P = .02$ ). Additionally, I-CoW anatomy was associated with a higher 90-day mortality than C-CoW (OR [95% CI], 2.58 [1.01–6.60];  $P = .04$ ) and NI-CoW (OR [95% CI], 1.89 [1.13–3.18];  $P = .01$ ).

**CONCLUSIONS:** CoW variants are associated with functional and mortality outcomes in patients treated with EVT for anterior circulation large vessel occlusion. Further research is needed to determine how CoW vessel anatomy may impact clinical assessment, triage, and treatment in acute ischemic stroke.

**ABBREVIATIONS:** ACA = anterior cerebral artery; ACom = anterior communicating artery; AIS = acute ischemic stroke; CoW = circle of Willis; C-CoW = complete circle of Willis; EVT = endovascular thrombectomy; I-CoW = isolating incomplete circle of Willis; LVO = large vessel occlusion; mTICI = modified treatment in cerebral infarction; NI-CoW = nonisolating incomplete circle of Willis; PCA = posterior cerebral artery; PCom = posterior communicating artery; Tan CS = Tan collateral scores

Cerebral collateral circulation is an essential variable for predicting functional outcomes in acute ischemic stroke (AIS) due to large vessel occlusion (LVO). Collateral blood flow to the ischemic penumbra is a positive predictor of smaller infarct growth, a lower risk of hemorrhagic transformation, and extended therapeutic windows.<sup>1,2</sup>

The circle of Willis (CoW) is a major source of collateral blood flow in the brain.<sup>3</sup> The anterior communicating artery (ACom) and posterior communicating artery (PCom) of the

CoW provide collateral flow between the left and right hemispheres and the anterior and posterior circulation, respectively.<sup>4</sup> Variance in these, as well as other constituents of the CoW, are common, with estimates indicating 54%–83% of the population having an incomplete CoW due to hypoplasia or aplasia in at least 1 vessel comprising the CoW.<sup>5</sup> For example, estimates from cadaveric studies regarding hypoplasia of the ACom range from 9%–30%, with total aplasia of the ACom in nearly 2% of the general population.<sup>6</sup> Similarly, hypoplasia or aplasia of the PCom occurs in approximately 68% of the population.<sup>5</sup>

Received August 14, 2024; accepted after revision December 19.

From the Departments of Radiology (A.S., M.K.O., A.S.V., D.M., J.V., L.L.-L.W.), and Neurology and Rehabilitation Medicine (S.L.D., Y.N.A.), University of Cincinnati Medical Center, Cincinnati, Ohio; University of Cincinnati College of Medicine (B.Z.), Cincinnati, Ohio; Division of Biostatistics (B.Z.), Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; Department of Diagnostic Radiology (S.V.), University of Pennsylvania, Philadelphia, Pennsylvania; Department of Diagnostic Radiology (A.A.), Henry Ford Medical Center, Ann Arbor, Michigan; and Department of Diagnostic Radiology (T.T.), Oregon Health and Science University, Portland, Oregon.

This work was supported by the NINDS NeuroNEXT grant and by American Heart Association Grant <https://doi.org/10.58275/AHA.24CDA1268532.pc.gr.193580>

Aakanksha Sriwastwa and Michael K. Oswald are co-first authors of this manuscript. Yasmin N. Aziz and Lily Li-Li Wang are co-last authors of this manuscript.

Please address correspondence to Yasmin N. Aziz, MD, Department of Neurology and Rehabilitation Medicine, University of Cincinnati Medical Center, 3188 Bellevue Ave, ML0761, Cincinnati, OH 45219; e-mail: [azizyn@ucmail.uc.edu](mailto:azizyn@ucmail.uc.edu); @ucstroke team

 Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8643>

## SUMMARY

**PREVIOUS LITERATURE:** Knowledge about how circle of Willis (CoW) collaterals affect large vessel stroke outcomes is lacking. Some studies have observed functional and mortality benefits in patients having a complete CoW versus an incomplete CoW. Others have found no such benefit or have only observed benefit in a select patient population (ie, internal carotid artery occlusions) or with specific vessels that comprise the CoW (for example, posterior communicating artery). We aimed to analyze the impact of variant CoW anatomy relative to the occluded anterior large vessel on stroke outcomes in patients treated successfully with mechanical thrombectomy.

**KEY FINDINGS:** Analysis of variant CoW anatomy in 465 patients with successfully revascularized anterior LVO demonstrated significant association of isolating CoW variants with higher 90-day mortality rates ( $P = .01$ ) and poorer functional outcome ( $P = .02$ ) compared with nonisolating CoW variants. Higher 90-day mortality was also noted when compared with the patients with complete CoW ( $P = .04$ ).

**KNOWLEDGE ADVANCEMENT:** In patients with successfully treated anterior circulation LVO, isolating CoW variant anatomy negatively impacts the 90-day functional and mortality outcome. This critical observation, if replicated in larger multicenter studies, may impact triage and treatment in acute ischemic stroke.

The impact of CoW variance on anterior LVO stroke is not completely understood. This is likely due to the mitigating effect of leptomeningeal collaterals, which are a second important source of collateralization during AIS. The leptomeningeal arteries direct blood via retrograde flow, whereas the CoW vessels supply antegrade blood flow.<sup>7</sup> An incomplete CoW variant has been linked to higher odds of developing an AIS.<sup>8-10</sup> However, an incomplete CoW does not necessarily result in poor collateral flow in the affected hemisphere. While some studies have found that CoW completeness is associated with better 90-day functional outcomes, others have not.<sup>11-15</sup> These studies are limited by small sample size and subsequent inability to control for leptomeningeal collaterals. Furthermore, the effect of hypoplastic or absent CoW arteries isolating the anterior LVO-affected hemisphere from the rest of the intracranial circulation (posterior or contralateral anterior) has not been analyzed.

We aimed to determine the relationship between 3 anatomically distinct CoW subtypes and 90-day functional outcomes in patients with AIS after successful endovascular thrombectomy (EVT). The CoW subtypes were classified based on their ability to provide collateralization from the posterior and/or contralateral cerebral hemispheres. We hypothesized that CoW subtypes that do not isolate the occluded anterior large vessel and the affected cerebral hemisphere from the posterior and contralateral circulation would be correlated with better functional and mortality outcomes. This paper follows the Strengthening the Reporting of Observational Studies in Epidemiology Checklist reporting guidelines.

## MATERIALS AND METHODS

### Study Population

We retrospectively analyzed prospectively collected data from an institutional registry of all consecutive patients with LVO treated at a comprehensive tertiary stroke center between May 2016 and November 2023. Patients who presented with AIS stroke due to occlusion of the ICA or the M1 segment of the MCA and underwent successful EVT (defined as modified TICI [mTICI] score 2b or greater) within 24 hours of symptom onset were included.<sup>16</sup> Patients presenting beyond 24 hours of symptom onset or last

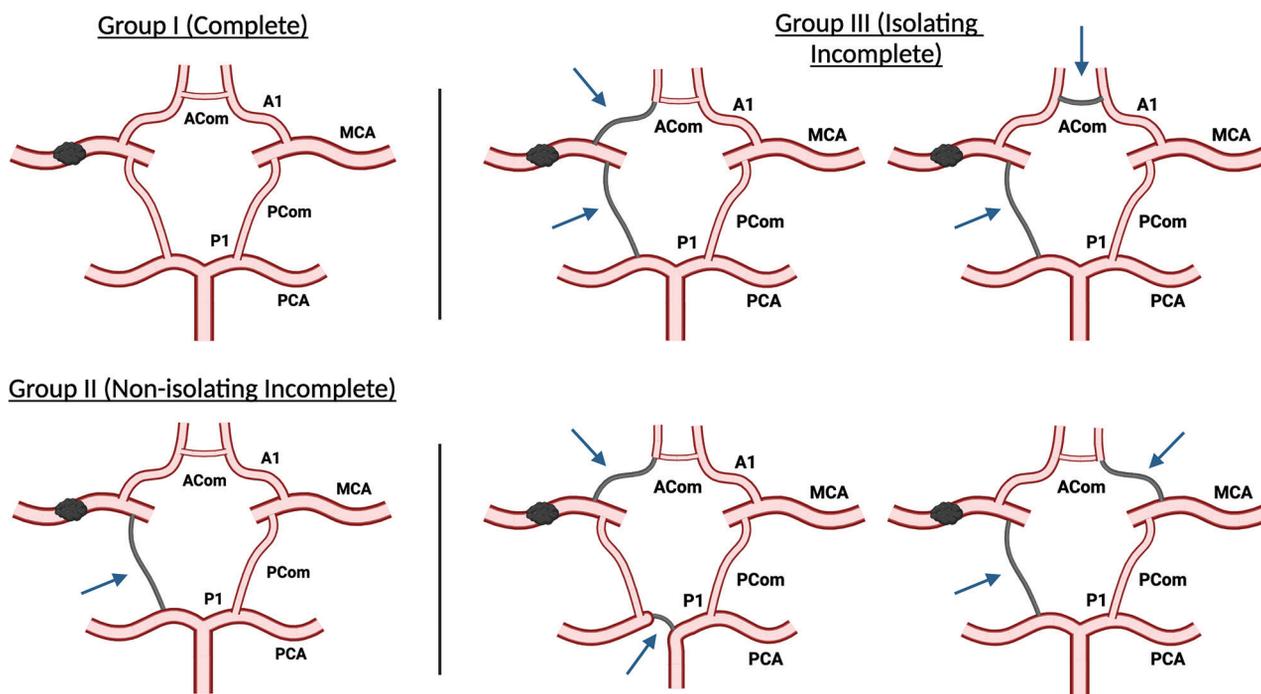
known well, those with tandem occlusions, or with mTICI 2a or lower post-EVT score, were excluded. The local LVO registry was approved under umbrella protocol #2016-6858 by the institutional review board. The requirement for informed consent was waived because of the retrospective nature of the study.

### Covariates and Study Outcomes

Variables including demographics (age, sex, race), medical history (hypertension, hyperlipidemia, congestive heart failure, diabetes mellitus, atrial fibrillation, smoking, prior history of stroke), and clinical characteristics (NIHSS score, point of care blood glucose [mg/dL], IV thrombolysis administration, and time to revascularization [minutes]) were obtained from the registry. NIHSS scores were recorded at the time of the first encounter. The primary outcome of this study was functional outcome, defined by the 90-day mRS score after initial stroke symptoms or time of last known well.<sup>17</sup> The 90-day mRS was collected by stroke study nurses. The score was dichotomized into good functional outcome (mRS 0-2; normal to mild disability with most activities of daily living being independently maintained) and poor functional outcome (mRS 3-6; moderate to severe disability, or death).

### Imaging Variables

CT head reports had ASPECTS for 395 of 465 cases (85.2%) included in the analysis. For the remaining cases, 2 senior radiology residents (J.V., D.M.) and 1 neuroradiology fellow (T.T.) reviewed the CT head to determine ASPECTS. A neuroradiology fellow with an additional fellowship in neuroradiology research (A.S.) independently adjudicated all cases for accuracy. CTA head imaging was reviewed to determine the site of vascular occlusion, CoW anatomy, and Tan collateral scores (Tan CS) for all 465 cases.<sup>18</sup> The radiology trainees reviewed the scans independently, and any disagreements in their assessments were re-evaluated. Discrepancies were handled by a board-certified neuroradiologist (L.W.). All radiologists were blinded to patient outcomes. The CoW anatomy was trichotomized by using baseline CTA of the head as follows (illustrated in Fig 1 and representative examples on CTA in the Supplemental Data):



**FIG 1.** Three distinct anatomic groups of CoW vessels. Group I (C-CoW): all 7 CoW arteries are normally developed, as illustrated. Group II (NI-CoW): incomplete but functionally NI-CoW variants that allow communication from the contralateral anterior or posterior circulations to the occluded vessel. There are many specific variants that apply to this category. Here, we show just 1 for illustrative purposes. Group III (I-CoW): comprises cases with functionally I-CoW variants that preclude communication with the occluded vessel from contralateral anterior circulation and posterior circulation. Four examples are depicted here for illustrative purposes. The arrows in the figures represent hypoplastic or absent CoW arteries. Illustration created by using software from BioRender.com. A1 = A1 segment-ACA; P1 = P1 segment-PCA.

- Group I, complete circle of Willis (C-CoW): All 7 arteries that comprise the CoW are robust and well-opacified by contrast. These include the A1 segments of the anterior cerebral artery (ACA), P1 segments of the posterior cerebral artery (PCA), PCom arteries, and ACom arteries.
- Group II, nonisolating incomplete circle of Willis (NI-CoW): Variants created by absent or hypoplastic (<1 mm in caliber) CoW vessels that have the potential to limit blood flow but do not isolate the occluded artery and affected cerebral hemisphere from the posterior circulation or the contralateral anterior circulation.
- Group III, isolating incomplete circle of Willis (I-CoW): Variants created by absent or hypoplastic CoW vessels that isolate the occluded artery and affect the cerebral hemisphere from the posterior circulation and contralateral anterior circulation.

Tan CS in the affected cerebral hemisphere was assessed on baseline CTA brain as follows: 0, absent collateral supply; 1, supply filling >0% but ≤50%; 2, supply filling >50% but <100%; and 3, 100% collateral filling of the occluded MCA territory.<sup>18</sup> ASPECTS, assessed for each patient on baseline CT head, and the site of vascular occlusion, assessed on CTA, were obtained from clinical neuroradiology attending interpretation if present in the electronic medical record. If not documented, the ASPECTS were read by radiology fellows (A.S. and T.T.).

The final mTICI score was obtained from the LVO registry and abstracted from the procedure notes by the neurointerventionalist. mTICI scores were classified as follows: grade 0, no perfusion; grade 1, antegrade perfusion past the initial occlusion but

limited distal branch filling with little or slow distal perfusion; grade 2a, antegrade perfusion of less than one-half of the occluded target artery territory; grade 2b, antegrade perfusion of more than one-half of the occluded target artery territory; grade 2c, near complete antegrade perfusion except for slow flow or distal emboli in a few distal cortical vessels; grade 3, complete antegrade perfusion of the occluded target artery territory.<sup>19</sup> Patients who achieved mTICI 2b or greater postrevascularization were included in the analysis.

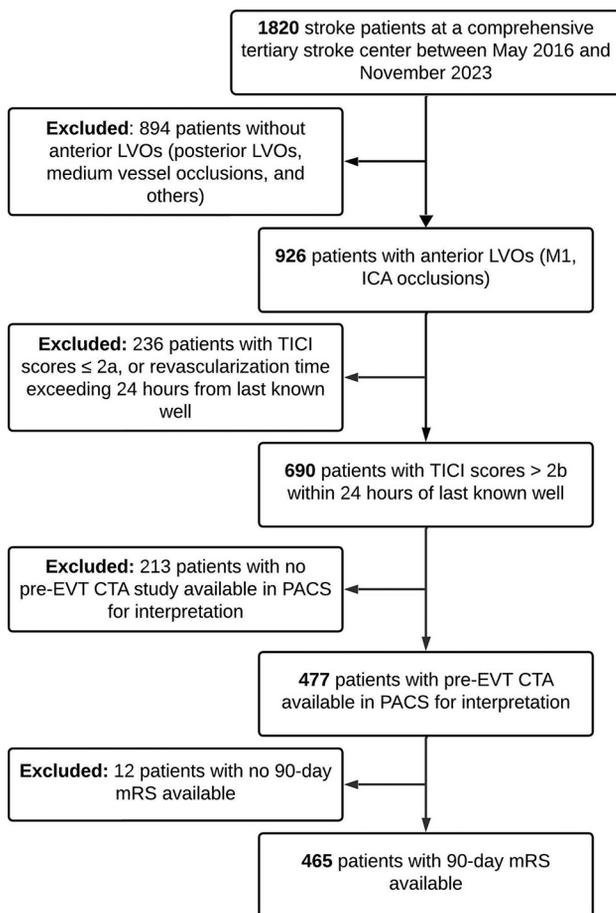
### Statistical Analysis

Descriptive analyses of patient demographics and clinical characteristics are reported as mean ± SD for age and count (percentage) for sex and medical history. The normality of the distribution of data points was assessed by using the Kolmogorov-Smirnov test. The  $\chi^2$  test and univariate logistic regression models were used to evaluate the association of CoW groups with functional and mortality outcomes. Predetermined covariates associated with stroke outcomes (age, sex, baseline ASPECTS, admission NIHSS, blood glucose levels, IV thrombolytic status, symptom onset to revascularization time, and Tan CS) were then fitted into multivariable logistic regression models along with the CoW groups.<sup>20-25</sup> All analyses were performed by using SAS Version 9.4 (SAS Institute). Statistical significance was set at  $P < .05$ .

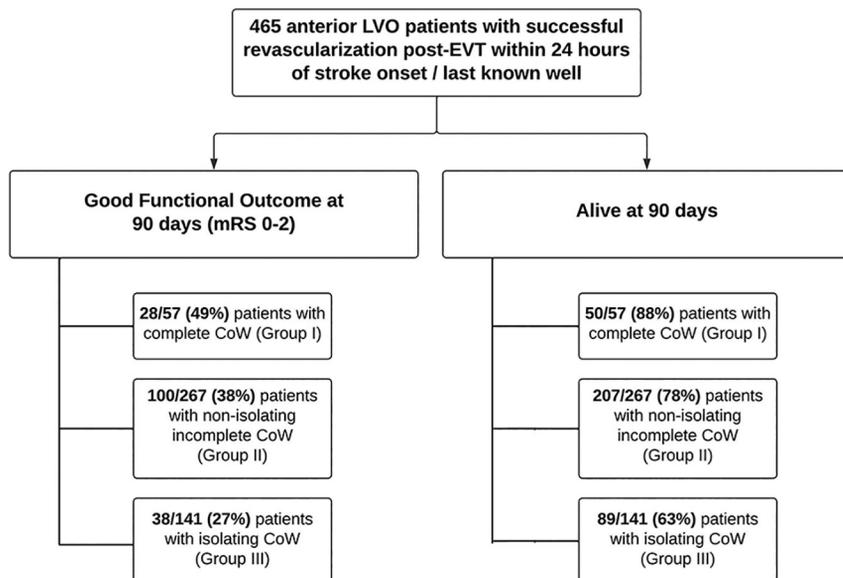
### RESULTS

Among the 1820 patients with AIS admitted to the comprehensive stroke center between May 2016 and November 2023, 465

met all inclusion criteria without any exclusion criteria (Fig 2). The mean age (SD) of the patients was 67.0 ( $\pm 15.3$ ) years; 51% were men, and 17% were black. All collected demographic, clinical, and imaging characteristic data are provided in the Supplemental



**FIG 2.** Flowchart outlining the steps of patient selection for the study.



**FIG 3.** Functional outcome and mortality at 90 days among the 3 CoW groups.

Data. A total of 57 (12%) patients demonstrated C-CoW, 267 (57%) had NI-CoW, and 141 (30%) had I-CoW. Good 90-day functional outcomes were observed in 166 (36%) patients and poor outcomes in 299 (64%) patients.

The percent rates of good functional outcomes and mortality among the 3 distinct CoW groups are shown in Fig 3. Individual mRS scores were analyzed among the 3 CoW groups, and a percentage-based trend toward worse outcomes and increased mortality was observed in the I-CoW group compared with the NI-CoW and C-CoW groups (Fig 4). The  $\chi^2$  test demonstrated a significant difference in the 90-day mRS ( $P = .007$ ) and mortality ( $P < .001$ ) between the 3 CoW groups. Univariate logistic regression analysis demonstrated a significant association between I-CoW and poor functional outcomes when compared with the C-CoW (OR [95% CI], 2.64 [1.40–5.00];  $P = .003$ ) and NI-CoW (OR [95% CI], 1.63 [1.04–2.55];  $P = .03$ ) groups (Table).

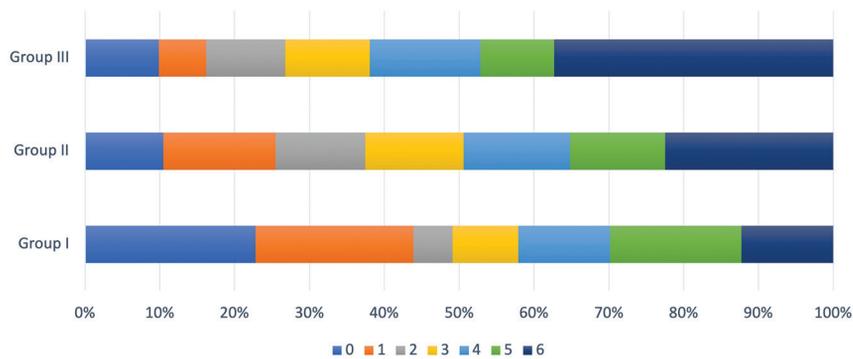
A multivariable logistic regression analysis was performed to assess the strength of the association after accounting for predetermined covariates (age, NIHSS, blood glucose, ASPECTS, Tan CS, and time to revascularization). A significant association of I-CoW with poor 90-day functional outcomes was again demonstrated when compared with NI-CoW (OR [95% CI], 1.83 [1.08–3.09];  $P = .02$ ) (Table). However, there was no significant difference in functional outcomes between the I-CoW and C-CoW groups (OR [95% CI], 1.87 [0.89–3.92];  $P = .10$ ).

Multivariable logistic regression models also demonstrated a significant association between I-CoW and higher 90-day mortality compared with C-CoW (OR [95% CI], 2.58 [1.01, 6.60];  $P = .04$ ) and NI-CoW (OR [95% CI], 1.89 [1.13, 3.18];  $P = .01$ ) (Table). Additionally, a comparison of the absence of residual deficit (mRS 0) among the 3 CoW groups at day 90 poststroke revealed a significant difference ( $P = .01$ ) by using  $\chi^2$  analysis.

We performed a post hoc analysis after adding intravenous thrombolysis administration as a covariate in the existing logistic regression model. Intravenous thrombolysis did not attain significance when compared with 90-day functional outcome ( $P = .35$ ) and trended toward significance when compared with 90-day mortality ( $P = .06$ ). The analysis demonstrated significant association of the I-CoW group (group III) with worse 90-day mortality compared with both the NI-CoW (OR [95% CI]: 1.93 [1.18, 3.16],  $P = .01$ ) and C-CoW (OR [95% CI]: 2.68 [1.06, 6.78],  $P = .04$ ) groups. The association of the I-CoW group with 90-day functional outcome, however, became borderline insignificant when compared with the NI-CoW group (OR [95% CI]: 1.93 [0.95, 2.53],  $P = .08$ ) and insignificant when compared with the C-CoW group (OR [95% CI]: 1.55 [0.76, 3.18],  $P = .23$ ).

Analysis of the distribution of hypoplastic and absent CoW arteries in the incomplete CoW groups (I-CoW and NI-CoW) showed that the PCom arteries

**Circle of Willis groups and 90-day functional outcome in each group indicated by modified Rankin Scores (0-6)**



**FIG 4.** Graph demonstrating 90-day poststroke functional outcome indicated by mRS (0–6) in the 3 CoW groups, where 0 indicates no residual deficit and 6 indicates death.

**Multivariable logistic regression model assessing the association of CoW groups and other covariates with 90-day functional and mortality outcomes (good, mRS 0–2, versus poor, mRS 3–6)**

	Functional Outcome	Functional Outcome	Mortality Outcome	Mortality Outcome
	OR (95% CI)	P Value	OR (95% CI)	P Value
C-CoW versus I-CoW	1.87 (0.89, 3.92)	.10	2.58 (1.01, 6.60)	<b>.04</b>
NI-CoW versus I-CoW	1.83 (1.08, 3.09)	<b>.02</b>	1.89 (1.13, 3.18)	<b>.01</b>
Age	0.97 (0.96, 0.99)	< <b>.001</b>	0.95 (0.94, 0.97)	< <b>.001</b>
Sex	0.88 (0.56, 1.38)	.58	1.01 (0.62, 1.66)	.95
Baseline ASPECTS	1.18 (1.04, 1.34)	<b>.01</b>	1.17 (1.03, 1.33)	<b>.01</b>
NIHSS at admission	0.92 (0.89, 0.95)	< <b>.001</b>	0.94 (0.91, 0.98)	<b>.001</b>
Tan CS	2.50 (1.52, 4.11)	< <b>.001</b>	1.84 (1.11, 3.04)	<b>.01</b>
Revascularization time (every 10-minute increase)	0.99 (0.98, 1.00)	<b>.002</b>	1.00 (0.99, 1.01)	.87

**Note:**—Bold text represents statistical significance achieved.

were hypoplastic or absent in most cases compared with other CoW arteries (Supplemental Data). Specifically, the PComs were hypoplastic or absent in 90% (127/141) of the I-CoW group and 85% (226/267) of the NI-CoW group, although this difference was not significant ( $P = .34$ ). In contrast, AComs were hypoplastic or absent in 85% (119/141) of the I-CoW group and 14% (37/267) of the NI-CoW group, showing a significant difference between the 2 groups ( $P < .001$ ). These findings suggest that the ACom could be a greater determinant of collateral flow than the more commonly absent or hypoplastic PCom. The A1 segment of the ACA was hypoplastic or absent in the fewest cases, with 12% in I-CoW and 1.8% in NI-CoW ( $P < .001$ ).

**DISCUSSION**

In successfully treated anterior circulation patients with LVO, we found that I-CoW variants were associated with lower functional independence and higher mortality than NI-CoW variants at 90 days. I-CoW variants were also associated with a higher 90-day mortality than C-CoW variants. In the univariate logistic regression model, statistical significance was achieved in favor of good functional outcomes in the C-CoW group compared with the I-CoW group. Surprisingly, this association was not maintained in multivariate analysis. The reason for this is unclear

but may be due to the small sample size within the C-CoW subgroup ( $n = 57$ ).

Our findings are consistent with those of previous studies examining the role of CoW variants and their effect on functional outcomes in patients with AIS. For example, 1 study found that the odds of good 90-day functional outcomes (mRS 0–2) were nearly 12 times higher in patients with an ACom or PCom. When both the ACom and PCom were present, the same study found the odds of a good functional outcome to be 29 times higher. They also concluded that AComs had a higher impact on stroke outcomes.<sup>26</sup> Another study demonstrated that the presence of PCom is an independent predictor of survival.<sup>27</sup> Finally, a third study concluded that the presence of an ACom in ICA occlusions leads to more favorable leptomeningeal collateralization.<sup>28</sup> Our study is in alignment with these results as the analysis of AComs demonstrated a higher frequency (85% versus 14%) and significant association ( $P < .001$ ) of absent or hypoplastic ACom with isolated CoW compared with nonisolated CoW, further underscoring the importance of ACom in collateralization.

However, there have been a few conflicting reports on the impact of specific CoW constructs on stroke outcomes. Some of these studies, like ours, were exclusively performed in patients undergoing EVT.<sup>26,28-30</sup> Others included patients without consideration of intervention.<sup>9,31</sup> Irrespective of the inclusion criteria, most of these studies determined a significant association between CoW variants and severity of stroke, but to varying extents and qualifying conditions that may oppose our reported findings in this study. A 2023 systematic review of 11 studies determined that the CoW plays a crucial role in stroke outcomes when distal ICA occlusions are present, yet plays no role in M1 occlusions, citing a study performed by Westphal et al<sup>29</sup> in 2021.<sup>15</sup> However, this study acknowledges a trend toward better mortality outcomes in patients with a complete CoW, and it is possible that the sample size of the smaller subgroups of the CoW variants statistically limited the study. The overall conclusion of the systematic review agrees with our paper that the integrity of the CoW is important for stroke prognosis.<sup>15</sup> Another recent study evaluating stroke outcomes in 182 patients found that an incomplete CoW was not associated with functional outcomes after successful endovascular therapy.<sup>14</sup> Of note is that this study was limited by a smaller sample size compared with our investigation, and ACom patency was not included in the analysis due to concerns regarding its reliable detection on their imaging technique of choice (MRA). Despite this, the authors estimate a significant association may be reached in cohorts exceeding 3000 individuals.

Many studies concerning CoW anatomy and its correlation with long-term functional outcomes have focused on the presence of communicating arteries only or have analyzed the effect of incomplete versus complete CoW variants.<sup>9,13,14,31,32</sup> However, we could only find 1 previous study that analyzed the impact of the functional characterization of CoWs, in patients who had undergone EVT, relative to the occluded vessel and its associated vascular territory.<sup>28</sup> Our investigation is unique to the literature, as it provides a more comprehensive assessment of CoW anatomy to account for the potential for collateral blood flow beyond 1 vessel and includes a sample size 3 times larger than the most similarly designed study.<sup>28</sup> Furthermore, we classified these variants into 3 subgroups to better understand the cumulative effects of specific variants on outcomes. The fact that variants that isolate the ischemia-affected cerebral parenchyma from the posterior and contralateral anterior circulations result in worse outcomes emphasizes the mitigating influence of these variants on ischemia.

For the CoW anatomy to be truly isolated, 1 or more vessels from both the anterior CoW (comprising ACom and left and right A1-ACA) and posterior CoW (comprising left and right PCom and P1-PCA) need to be hypoplastic or absent. This anatomy was observed in nearly one-third of the patients in our study population, emphasizing that it is not uncommon, further underscoring the importance of CoW anatomy in stroke outcomes in a large proportion of patients despite the best efforts at revascularization and penumbra salvage. Additionally, the study findings suggest that the ACom is a greater determinant of collateral flow than the more commonly hypoplastic or absent PCom. This observed effect of anterior CoW variants on collateral flow has been seen in another study before ours, albeit concerning the A1 segment of the ACA in the other study.<sup>33</sup>

Importantly, the interdependence of CoW anatomy and Tan CS in determining leptomeningeal collateralization, and the resultant protective effect in ischemia, has not been well-analyzed in previous studies. In our patient cohort, a significant correlation was demonstrated between good (2–3) versus poor (0–1) Tan CS and the CoW groups on pair-wise comparison. Furthermore, in the multivariable logistic regression analysis, Tan CS were significantly associated with both 90-day functional and mortality outcomes after controlling for other covariates, including CoW groups. These results underscore the combined role that these 2 important determinants of leptomeningeal collateralization play in mitigating the harmful effects of vessel occlusion.

The post hoc analysis performed with intravenous thrombolysis as an additional covariate in the existing multivariable logistic regression model demonstrated significant association of CoW anatomy with 90-day mortality. However, it failed to demonstrate a significant association with the functional outcome. This lack of association was most likely because the study was not sufficiently powered to incorporate additional covariates. Larger studies are required to validate these results.

Our study is limited by its retrospective design, lack of information regarding premorbid mRS, time to needle, final infarct volume, anatomic or other physiologic circumstances that could affect collateral blood flow (ie, severe hypotension or heart failure), and dichotomized outcome scoring of functional outcomes.

Further, we did not include CTP imaging data because it is only performed routinely in a minority of patients before EVT. Finally, our rate of C-CoW is smaller than the rates reported in the literature, likely due to the numbers being derived from the general population and nonstroke patients. Our study is strengthened by a unique classification of CoW variants designed to account for potential stepwise decreases in collateral blood flow and a relatively large sample size that allowed for control of the leptomeningeal collateral score in outcome correlation.

The American Heart Association guidelines state that it may be reasonable to incorporate collateral flow status in EVT eligibility decision-making.<sup>34</sup> Considering the established relationship between CoW anatomy and collateral capacity, CoW anatomy may be a helpful consideration when predicting the outcomes of patients with AIS. Additionally, targeted analysis evaluating their impact on large core infarcts needs to be performed.

## CONCLUSIONS

CoW variants are associated with functional and mortality outcomes in patients treated with EVT for anterior large vessel occlusion. Further research is needed to determine how CoW vessel anatomy may impact the clinical assessment, triage, and treatment of AIS.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Vagal A, Aviv R, Sucharew H, et al. **Collateral clock is more important than time clock for tissue fate.** *Stroke* 2018;49:2102–07 CrossRef Medline
2. Bang OY, Goyal M, Liebeskind DS. **Collateral circulation in ischemic stroke: assessment tools and therapeutic strategies.** *Stroke J Cereb Circ* 2015;46:3302–09 CrossRef Medline
3. Hartkamp MJ, van Der Grond J, van Everdingen KJ, et al. **Circle of Willis collateral flow investigated by magnetic resonance angiography.** *Stroke* 1999;30:2671–78 CrossRef Medline
4. Powers WJ. **Cerebral hemodynamics in ischemic cerebrovascular disease.** *Ann Neurol* 1991;29:231–40 CrossRef Medline
5. Jones JD, Castanho P, Bazira P, et al. **Anatomical variations of the circle of Willis and their prevalence, with a focus on the posterior communicating artery: a literature review and meta-analysis.** *Clin Anat N Anat* 2021;34:978–90 CrossRef Medline
6. Krzyżewski RM, Tomaszewski KA, Kochana M, et al. **Anatomical variations of the anterior communicating artery complex: gender relationship.** *Surg Radiology Anat* 2015;37:81–86 CrossRef Medline
7. Menon BK, Smith EE, Modi J, et al. **Regional leptomeningeal score on CT angiography predicts clinical and imaging outcomes in patients with acute anterior circulation occlusions.** *AJNR Am J Neuroradiol* 2011;32:1640–45 CrossRef Medline
8. Wang H, Shen L, Zhao C, et al. **The incomplete circle of Willis is associated with vulnerable intracranial plaque features and acute ischemic stroke.** *J Cardiovasc Magn Reson* 2023;25:23 CrossRef Medline
9. De Caro J, Ciacciarelli A, Tessitore A, et al. **Variants of the circle of Willis in ischemic stroke patients.** *J Neurol* 2021;268:3799–807 CrossRef Medline
10. Oumer M, Alemayehu M, Muche A. **Association between circle of Willis and ischemic stroke: a systematic review and meta-analysis.** *BMC Neurosci* 2021;22:3 CrossRef Medline
11. Rozeman A, Hund H, Boiten J, et al. **Circle of Willis variation and outcome after intra-arterial treatment.** *BMJ Neurol Open* 2022;4:e000340 CrossRef Medline

12. Zhao H, Wang B, Xu G, et al. **Collateral grade of the Willis' circle predicts outcomes of acute intracranial internal carotid artery occlusion before thrombectomy.** *Brain Behav* 2019;9:e01452 CrossRef Medline
13. Lin E, Kamel H, Gupta A, et al. **Incomplete circle of Willis variants and stroke outcome.** *Eur J Radiol* 2022;153:110383 CrossRef Medline
14. Seifert-Held T, Eberhard K, Christensen S, et al. **Circle of Willis variants are not associated with thrombectomy outcomes.** *Stroke Vasc Neurol* 2021;6:310–13 CrossRef Medline
15. Širvinskas A, Lengvenis G, Ledas G, et al. **Circle of Willis configuration and thrombus localization impact on ischemic stroke patient outcomes: a systematic review.** *Medicina (Kaunas)* 2023;59:2115 CrossRef Medline
16. Wang R, Huang J, Mohseni A, et al. **Predictors of mTICI 2c/3 over 2b in patients successfully recanalized with mechanical thrombectomy.** *Ann Clin Transl Neurol* 2024;11:89–95 CrossRef Medline
17. Banks JL, Marotta CA. **Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials.** *Stroke* 2007;38:1091–96 CrossRef Medline
18. Tan JC, Dillon WP, Liu S, et al. **Systematic comparison of perfusion-CT and CT-angiography in acute stroke patients.** *Ann Neurol* 2007;61:533–43 CrossRef Medline
19. LeCouffe NE, Kappelhof M, Treurniet KM, et al.; MR CLEAN Registry Investigators. **2B, 2C, or 3: what should be the angiographic target for endovascular treatment in ischemic stroke?** *Stroke* 2020;51:1790–96 CrossRef Medline
20. Fortunati V, Su J, Wolff L, et al. **Siamese model for collateral score prediction from computed tomography angiography images in acute ischemic stroke.** *Front Neuroimaging* 2023;2:1239703 CrossRef Medline
21. Adams HP, Davis PH, Leira EC, et al. **Baseline NIH Stroke Scale score strongly predicts outcome after stroke.** *Neurology* 1999;53:126–31 CrossRef Medline
22. Bagg S, Pombo AP, Hopman W. **Effect of age on functional outcomes after stroke rehabilitation.** *Stroke* 2002;33:179–85 CrossRef Medline
23. Risitano A, Toni D. **Time is brain: timing of revascularization of brain arteries in stroke.** *Eur Heart J Suppl* 2020;22:L155–59 CrossRef Medline
24. Ospel JM, Kappelhof M, Kashani N, et al.; HERMES Collaborators. **Effect of age and baseline ASPECTS on outcomes in large-vessel occlusion stroke: results from the HERMES collaboration.** *J Neurointerv Surg* 2021;13:790–93 CrossRef Medline
25. Hiraga A. **Gender differences and stroke outcomes.** *Neuroepidemiology* 2017;48:61–62 CrossRef Medline
26. Sablić S, Dolić K, Kraljević I, et al. **The presence of communicating arteries in the Circle of Willis is associated with higher rate of functional recovery after anterior circulation ischemic stroke.** *Biomedicines* 2023;11:3008 CrossRef Medline
27. Sadeh-Gonik U, Budylev A, Krivitzky D, et al. **Circle of Willis integrity in acute middle cerebral artery occlusion: does the posterior communicating artery matter?** *J Neurointerv Surg* 2024;16:801–08 CrossRef Medline
28. Millesi K, Mutzenbach JS, Killer-Oberpfalzer M, et al. **Influence of the circle of Willis on leptomeningeal collateral flow in anterior circulation occlusive stroke: friend or foe?** *J Neurol Sci* 2019;396:69–75 CrossRef Medline
29. Westphal LP, Lohaus N, Winklhofer S, et al. **Circle of Willis variants and their association with outcome in patients with middle cerebral artery-M1-occlusion stroke.** *Eur J Neurol* 2021;28:3682–91 CrossRef Medline
30. Lee AY, Campbell JP, Hwang TS, et al.; American Academy of Ophthalmology. **Recommendations for standardization of images in ophthalmology.** *Ophthalmology* 2021;128:969–70 CrossRef Medline
31. Zhou H, Sun J, Ji X, et al. **Correlation between the integrity of the circle of Willis and the severity of initial noncardiac cerebral infarction and clinical prognosis.** *Medicine (Baltimore)* 2016;95:e2892 CrossRef Medline
32. Shaban A, Albright KC, Boehme AK, et al. **Circle of Willis variants: fetal PCA.** *Stroke Res Treat* 2013;2013:105937 CrossRef Medline
33. Fischer S, Goertz L, Weyland CS, et al. **Functional aplasia of the contralateral A1 segment influences clinical outcome in patients with occlusion of the distal internal carotid artery.** *J Clin Med* 2022;11:1293 CrossRef Medline
34. Powers WJ, Rabinstein AA, Ackerson T, et al.; American Heart Association Stroke Council. **2018 Guidelines for the Early Management of Patients With Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association.** *Stroke* 2018;49:e46–110 CrossRef Medline

# Effect of SARS-CoV-2 Infection on Endovascular Thrombectomy Outcomes—Data from the Florida Stroke Registry

Hayes B. Fountain, Ian Ramsay, Ruijie Yin, Ahmed Abdelsalam, Michael Silva, David Z. Rose, Angus Jameson, Ying Hao, Ayham Alkhachroum, Carolina M. Gutierrez, Victor J. Del Brutto, Robert M. Starke, Tanja Rundek, Hannah Gardener, Jose G. Romano, and Negar Asdaghi, for the Florida Stroke Registry



## ABSTRACT

**BACKGROUND AND PURPOSE:** Endovascular thrombectomy outcomes are impacted by changes in stroke systems of care. During the pandemic, SARS-CoV-2 positive status had major implications on hospital arrival and treatment models of non-COVID-related hospital admissions. Using the Florida Stroke Registry, we compared the rates of in-hospital death and discharge outcomes of patients treated with endovascular thrombectomy who tested positive for SARS-CoV-2 infection during their hospitalization.

**MATERIALS AND METHODS:** Data from Get with the Guidelines—Stroke hospitals participating in the Florida Stroke Registry during the COVID pandemic from March 2020 to December 2022 were reviewed to identify endovascular thrombectomy patients with coding for SARS-CoV-2 testing during their hospital stay. Associations between SARS-CoV-2 status and favorable endovascular thrombectomy outcomes of mRS (0–2) at discharge, discharge to home or rehabilitation center, symptomatic intracerebral hemorrhage, in-hospital mortality, and independent ambulation at discharge were examined by using multivariate logistic regression modeling adjusting for demographics, vascular risk factors, and clinical characteristics. Temporal analyses were used to compare outcomes across the study period.

**RESULTS:** A total of 8184 patients underwent endovascular thrombectomy (median age 71.1 years, women 49.7%, mean NIHSS 14). Of these, 180 (2.20%) were SARS-CoV-2 positive. Compared with SARS-CoV-2 negative endovascular thrombectomy patients, those who tested positive were younger, more frequently men, but with comparable stroke severity at presentation. In multivariable analysis, adjusting for baseline differences and confounding variables, there was a 33% lower likelihood of being discharged to home/inpatient rehab (OR = 0.67, 95% CI: 0.49–0.93), 65% higher odds of in-hospital death (OR = 1.65, 95% CI: 1.06–2.58), as well as a 91% less chance of having a high mRS (>2) at discharge (OR = 0.15, 95% CI: 0.04–0.60) for patients with positive SARS-CoV-2 infection. However, a similar risk of symptomatic intracerebral hemorrhage was present compared with SARS-CoV-2 negative patients (OR = 0.97, 95% CI: 0.50–1.88). Temporal analysis of SARS-CoV-2 positive patients showed no significant differences.

**CONCLUSIONS:** In this large multicenter stroke registry, despite comparable clinical presentation and treatment timelines, SARS-CoV-2 positive status negatively impacted thrombectomy outcomes.

**ABBREVIATIONS:** AIS = acute ischemic stroke; EVT = endovascular thrombectomy; FSR = Florida Stroke Registry; LVO = large vessel occlusion; sICH = symptomatic intracerebral hemorrhage

Endovascular thrombectomy (EVT) is the standard of care for the treatment of select patients with acute stroke related to a large vessel.<sup>1</sup> Additionally, it is also known that post EVT care in a dedicated stroke unit or neurologic intensive care unit is superior to post-EVT care in other units.<sup>2,3</sup> These treatment

paradigms, or systems of care, are essential in ensuring optimal outcomes in patients receiving EVT for large vessel acute ischemic strokes (AIS).

The SARS-CoV-2 pandemic created unprecedented challenges to every aspect of health care, especially in systems of care and resource allocation.<sup>4</sup> These challenges were amplified in the setting of emergency care and time-sensitive treatments, including EVT for large vessel AIS. Though there has been research examining the relationship between SARS-CoV-2 and AIS, as well as AIS and EVT outcomes, there are limited data reports on outcomes following EVT in patients with AIS who are SARS-CoV-2 infection positive.<sup>5–12</sup>

The Florida Stroke Registry (FSR) is a statewide database that collects data on stroke care throughout the state. FSR provides a

Received September 4, 2024; accepted after revision January 8, 2025.

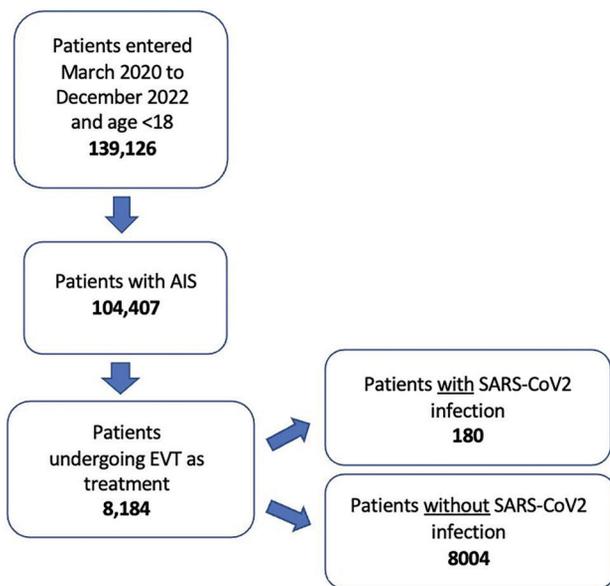
From the Department of Neurology (H.B.F., I.R., R.Y., A.A., M.S., Y.H., A.A., C.M.G., V.J.D.B., R.M.S., T.R., H.G., J.G.R., N.A.), Miller School of Medicine, University of Miami, Miami, Florida; and University of South Florida Morsani College of Medicine (D.Z.R., A.J.), Tampa, Florida.

Please address correspondence to Hayes B. Fountain, MD, BA, 1120 NW 14th St, Clinical Research Building, 13th floor, Miami, FL 33136; e-mail: hfountain@med.miami.edu



Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8673>



**FIG 1.** Details of study participant inclusion flow chart.

unique opportunity to examine the impact of SARS-CoV-2 infection status on the outcomes of EVT in patients with AIS. We compared in-hospital and discharge outcomes of endovascularly treated patients with versus without a SARS-CoV-2 positive infection status during their hospitalization.

## MATERIALS AND METHODS

Data from Get with the Guidelines–Stroke hospitals participating in the Florida Stroke Registry from March 2020 to December 2022 were reviewed. This study period was selected because, on March 9, 2020, the Florida Governor issued an executive order that declared a state of emergency for the entire state due to COVID-19.<sup>13–16</sup> Figure 1 describes the details of study inclusion. Patients who had an International Classification of Diseases, 10th revision (ICD-10) procedural code for EVT and coding for SARS-CoV-2 testing during their hospital stay were included. EVT was defined as the use of any form of mechanical thrombectomy device for the treatment of an acute intracranial occlusion with or without intra-arterial thrombolysis. These patients were categorized into SARS-CoV-2 positive or negative groups (Supplemental Data). Associations between SARS-CoV-2 status, clinical presentation, and post-EVT hospital outcomes were studied. Outcomes including mRS (0–2), discharge to home or to an acute rehabilitation center, independent ambulation at discharge, symptomatic intracerebral hemorrhage (sICH), and in-hospital mortality were examined by using multivariate logistic regression modeling adjusting for demographics, vascular risk factors, and clinical characteristics with generalized estimating equations. ORs with 95% CIs were calculated by using multivariate modeling with age, sex, race, insurance, smoker status, hypertension, diabetes, dyslipidemia, atrial fibrillation, history of coronary artery disease, previous stroke, arrival mode, IV thrombolysis, region, large vessel occlusion (LVO) location, admission NIHSS, period, onset to arrival, door to CT, door to needle, door to puncture, onset to puncture, onset to puncture <6 hours, onset to puncture

6 to 24 hours, onset to puncture >24 hours, and admitting service as covariates. Two different models were run to elucidate the most impactful covariates and minimize confounding. The first multivariate model used age, sex, insurance, race, EMS delivery mode, time variables, tPA administration, diabetes mellitus status, and diagnosis of atrial fibrillation. The second multivariate model used all previously mentioned variables and also the remaining covariates of smoker status, hypertension, hyperlipidemia, coronary artery disease, prior stroke, LVO location, arrival NIHSS score, admitting service, and admission care. Temporal analyses compared outcomes in SARS-CoV-2 positive patients across the years 2020–2022 by using 2020 as the index year and comparing against 2021 and 2022. Temporal analysis was completed on only the outcomes of good mRS at discharge, discharge to home or an acute rehabilitation center, sICH, and in-hospital mortality due to a low number of patients independently ambulating at baseline. For covariates with informative missingness, the Missing Indicator Method, which adds indicator variables to indicate the missing pattern, was used in the multivariate analysis to improve model performance.

## RESULTS

A total of 8184 patients underwent EVT during the study period, of which 180 (2.20%) were SARS-CoV-2 positive. Study participant characteristics and hospital characteristics are presented in Supplemental Data.

The patients with SARS-CoV-2 infection were more likely to be younger (67.2 years versus 71.1 years), be men (58.3% men versus 41.7% women), have increased onset to arrival time (229 minutes versus 161 minutes), and have increased onset to puncture time (379 minutes versus 291 minutes).

Overall, 80.78% of all EVT patients had an mRS >2 at discharge, 4670 (57.06%) were independently ambulating, and 4760 (58.16%) were discharged to home or a rehabilitation center. Additionally, overall, EVT patients, 447 (5.46%) had sICH, and 783 (9.57%) died in the hospital. Table 1 details in-hospital mortality, rate of sICH, and favorable discharge outcomes stratified based on SARS-CoV-2 status. As shown in Table 2, adjusting for baseline differences and confounding variables there was a 33% lower likelihood of being discharged to home/inpatient rehab (OR = 0.67, 95% CI: 0.49–0.93), 65% higher odds of in-hospital death (OR = 1.65, 95% CI: 1.06–2.58), as well as a 91% less chance of having a high mRS (>2) at discharge (OR = 0.15, 95% CI: 0.04–0.60) for patients with positive SARS-CoV-2 infection. However, a similar risk of symptomatic intracerebral hemorrhage was present compared with SARS-CoV-2 negative patients (OR = 0.97, 95% CI: 0.50–1.88).

Temporal analysis of year-to-year comparisons of outcomes by using 2020 as the index year showed no significant differences in odds of a good mRS at discharge (2020 versus 2021 OR 1.54; 95% CI: 0.89, 2.67 and 2020 versus 2022 OR 1.04; 95% CI: 0.47, 2.28), sICH (2020 versus 2021 OR 0.78; 95% CI: 0.54, 1.11 and 2020 versus 2022 OR 0.87; 95% CI: 0.59, 1.27), and in-hospital mortality (2020 versus 2021 OR 0.89; 95% CI: 0.75, 1.05 and 2020 versus 2022 OR 0.92; 95% CI: 0.74, 1.14), discharge to home or acute rehab center (2020 versus 2021 OR 0.95; 95% CI: 0.86, 1.05 and 2020 versus 2022 OR 1.02; 95% CI: 0.91, 1.14) in SARS-CoV-2

**Table 1: Raw data for investigated patient outcomes<sup>a</sup>**

Variable	Level	Overall	SARS-CoV-2? No	SARS-CoV-2? Yes	P Value
		n = 8184	n = 8004	n = 180	
Discharge location	1-Home	2894 (35.36%)	2839 (35.47%)	55 (30.56%)	.396
	2-Rehab	1866 (22.80%)	1827 (22.83%)	39 (21.67%)	
	3-Hosp	1059 (12.94%)	1031 (12.88%)	28 (15.56%)	
	4-SNF	1383 (16.90%)	1354 (16.92%)	29 (16.11%)	
	5-Other	199 (2.43%)	194 (2.42%)	5 (2.78%)	
	6-Died	783 (9.57%)	759 (9.48%)	24 (13.33%)	
Discharge ambulation	1-Able to ambulate independently	4670 (57.06%)	4555 (56.91%)	115 (63.89%)	.171
	2-Unable/with assistance	355 (4.34%)	348 (4.35%)	7 (3.89%)	
	3-ND/missing	3159 (38.60%)	3101 (38.74%)	58 (32.22%)	
Discharge mRS	High (>2)	1488 (80.78%)	1445 (80.41%)	43 (95.56%)	.011
	Low (0–2)	354 (19.22%)	352 (19.59%)	2 (4.44%)	
Symptomatic intracerebral hemorrhage <36 hrs	N	7737 (94.54%)	7566 (94.53%)	171 (95.00%)	.783
	Y	447 (5.46%)	438 (5.47%)	9 (5.00%)	
In-hospital mortality, %	Y	783 (9.57%)	759 (9.48%)	24 (13.33%)	.197
	N	7390 (90.30%)	7234 (90.38%)	156 (86.67%)	

Note:—Hosp indicates Hospital; SNF, Skilled Nursing Facility; ND, Not done/missing.

<sup>a</sup> Values are presented as "mean (% of total)" unless otherwise noted.

**Table 2: Outcomes of SARS-CoV-2 positive versus negative patients with AIS undergoing EVT<sup>a</sup>**

Outcome	Univariate ORs (95% CI)	Multivariate Model 1 ORs (95% CI)	Multivariate Model 2 ORs (95% CI)
SARS-CoV-2 positive versus negative EVT patients			
Favorable discharge location (home/rehabilitation vs other)	0.82 (0.616, 1.091), P = .1733	0.66 (0.48, 0.90), P = .0092	0.67 (0.49, 0.93), P = .0146
Good mRS at discharge (0–2)	0.608 (0.054, 6.779), P = .6857	0.17 (0.04, 0.62), P = .0075	0.15 (0.04, 0.60), P = .0071
sICH	0.956 (0.493, 1.854), P = .8951	1.00 (0.52, 1.94), P = .9904	0.97 (0.50, 1.88), P = .9321
In-hospital mortality	0.1735 (0.474, 1.128), P = .1569	1.58 (1.02, 2.45), P = .0408	1.65 (1.06, 2.58), P = .0274
Independent ambulation at discharge (restricted to patients who were independently ambulating at presentation)	0.989 (0.664, 1.472), P = .9562	0.92 (0.61, 1.39), P = .6899	0.95 (0.62, 1.45), P = .8097

<sup>a</sup> Multivariate model 1 adjusts for age, sex, race, insurance, diabetes, atrial fibrillation, arrival mode, IV thrombolysis, and period. Multivariate model 2 adjusts for all variables in Supplemental Data.

**Table 3: Temporal analysis of outcomes in SARS-CoV-2 positive patients to examine changes throughout the pandemic (2020–2022)**

Outcome	Period	aOR	95% CI
Favorable discharge location (home/rehabilitation vs other)	2020 vs 2021	0.95	(0.86, 1.05)
	2020 vs 2022	1.02	(0.91, 1.14)
Good mRS at discharge (0–2)	2020 vs 2021	1.54	(0.89, 2.67)
	2020 vs 2022	1.04	(0.47, 2.28)
sICH	2020 vs 2021	0.78	(0.54, 1.11)
	2020 vs 2022	0.87	(0.59, 1.27)
In-hospital mortality	2020 vs 2021	0.89	(0.75, 1.05)
	2020 vs 2022	0.92	(0.74, 1.14)

positive versus negative patients when comparing the years of 2020 to 2021 and 2020 to 2022 (Table 3).

## DISCUSSION

Despite younger age and similar timelines of presentation and neurologic severity, when controlling for confounding variables, EVT-treated patients who were SARS-CoV-2 positive had lower odds of achieving favorable discharge outcomes of discharge to home or rehabilitation center, good mRS (0–2) at discharge, and higher odds of in-hospital mortality.

The currently available studies on SARS-CoV-2 status and EVT outcomes for patients with AIS are equivocal about outcomes following EVT.<sup>5–12</sup> Many of the available studies are of smaller cohorts ( $n < 50$ ), and some are across many countries or regions, leading to the potential for differences in systems of care and treatment guidelines that may not represent true differences

in stroke outcomes between SARS-CoV-2 positive and negative patients. Our analysis of Get with the Guidelines–Stroke data across the state of Florida shows that patients with SARS-CoV-2 receiving EVT for LVO strokes are less likely to have a favorable discharge location (home/rehab versus others) and a good mRS (0–2) at discharge, while also having increased risk of in-hospital mortality. Although other groups have posited it that these unfavorable outcomes in patients with SARS-CoV-2 receiving EVT may be the result of the natural course of the SARS-CoV-2 disease, our group believes that there is another prominent factor involved in these differences: changes in systems of care and clinical treatment pathways that were instituted as the SARS-CoV-2 pandemic descended on the world.<sup>17–20</sup> These systemic changes may be reflected in multiple data points in this study, such as the increased time from onset to groin puncture (average 379 versus 291 minutes) shown in Supplemental Data. Increased times were

seen in onset to arrival, door to needle, and door to CT time as well. These increases in time may be reflective of the increased time needed to don personal protective equipment, such as respirators, N95 masks, and gowns for both EMS personnel and receiving nurses and physicians in the emergency department and operating rooms. Clinical treatment pathways were also altered as the pandemic began. For example, some institutions require intubation of patients in the emergency department before transportation to the endovascular suite for EVT. These patients also required isolation, especially early in the pandemic, when stopping the spread of the disease was of great importance as no vaccine was available and little was known about the disease. Additionally, some patients may have been more reluctant to seek acute care for stroke symptoms, leading to increased time from symptom onset to groin puncture. This is shown in our data with greater percentages of SARS-CoV-2 patients having onset to puncture times 6–24 hours or >24 hours rather than <6 hours. The reluctance of all people to obtain medical care during the pandemic is well-documented, even among those with acute stroke.<sup>20</sup> Fear of spreading the disease and decreased casual interactions that often lead to the discovery of a loved one with a stroke syndrome may have been reasons for these increased times between presentation and treatment in patients with SARS-CoV-2.

The impact of SARS-CoV-2 on the brain has been well-documented, spanning the investigation of cellular endothelial changes in cerebral blood vessels to acute ischemic events in the setting of SARS-CoV-2 predicting worse outcomes in patients with clinical encephalopathy.<sup>21–23</sup> These impacts on brain vasculature and the fact that CNS ischemic events as a cause of acute encephalopathy has been shown to be a predictor of poor outcomes present another explanation for worse outcomes among patients with SARS-CoV-2 undergoing EVT for large vessel AIS.

Another system-of-care factor that may have played a role in the observed differences in outcomes seen here is the admitting service and locations where care was administered. It has been shown that care in a dedicated neurologic unit (neuro ICU, stroke unit) independently improves the outcomes of EVT patients.<sup>2,3</sup> As the pandemic began, patients were being cared for in non-neurologically specific environments by non-neurologic providers to stop the spread of the disease by concentrating SARS-CoV-2 patients in one place, thereby limiting the exposure of other patients and staff. Another explanation for the observed differences in outcomes for SARS-CoV-2 positive patients in this study may be the differences in admitting service/care service; however, we have too large an amount of missing data to answer this question at present.

As this will not likely be the last pandemic, it is relevant and important to keep the SARS-CoV-2/COVID-19 impact on systems of care for AIS in mind. By doing this we can create resiliency in our systems of care to avoid these discrepancies in future outcomes.

### Limitations

The retrospective nature of this study is an inherent limitation and potential source of bias. Additionally, significant missing data in some of the variables in the FSR database are a potential source of confounding.

## CONCLUSIONS

In this large multicenter stroke registry, we found that despite similar risk factors, neurologic symptom severity, NIHSS scores, and in-hospital treatment timelines, SARS-CoV-2 positive patients were more likely to be younger, be men, have increased onset to arrival time, and have increased onset to puncture time. The major finding of our study was that increased treatment timelines for SARS-CoV-2-positive patients were associated with a lower likelihood of discharge to home or a rehabilitation center, a lower likelihood of mRS <2 at discharge, and a higher likelihood of in-hospital mortality. Temporal assessment showed similar results across the years of 2020, 2021, and 2022. These findings provide novel insight from a large database to add to the emerging literature examining outcomes of concurrent SARS-CoV-2 infection in the setting of EVT for target vessel occlusion AIS. Future studies should examine EVT outcomes by using a larger group of SARS-CoV-2 positive patients, as increased analytic power may provide further insights about the associations reported here.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. McCarthy DJ, Diaz A, Sheinberg DL, et al. **Long-term outcomes of mechanical thrombectomy for stroke: a meta-analysis.** *Scientific World Journal* 2019;2019:7403104 CrossRef Medline
2. Jørgensen HS, Nakayama H, Raaschou HO, et al. **The effect of a stroke unit: reductions in mortality, discharge rate to nursing home, length of hospital stay, and cost. A community-based study.** *Stroke* 1995;26:1178–82 CrossRef Medline
3. Langhorne P, Ramachandra S, Stroke Unit Trialists' Collaboration. **Organised inpatient (stroke unit) care for stroke: network meta-analysis.** *Cochrane Database Syst Rev* 2020;4:CD000197 CrossRef Medline
4. Moynihan R, Sanders S, Michaleff ZA, et al. **Impact of COVID-19 pandemic on utilisation of healthcare services: a systematic review.** *BMJ Open* 2021;11:e045343 CrossRef Medline
5. Zureigat H, Alhusban M, Cobia M. **Mechanical thrombectomy outcomes in COVID-19 patients with acute ischemic stroke: a narrative review.** *Neurologist* 2021;26:261–67 CrossRef Medline
6. Wang Z, Teng H, Wu X, et al. **Efficacy and safety of recanalization therapy for acute ischemic stroke with COVID-19: a systematic review and meta-analysis.** *Front Neurol* 2022;13:984135 CrossRef Medline
7. Al Kasab S, Almallouhi E, Alawieh A; STAR Collaborators, et al. **International experience of mechanical thrombectomy during the COVID-19 pandemic: insights from STAR and ENRG.** *J Neurointerv Surg* 2020;12:1039–44 CrossRef Medline
8. de Havenon A, Yaghi S, Mistry EA, et al. **Endovascular thrombectomy in acute ischemic stroke patients with COVID-19: prevalence, demographics, and outcomes.** *J Neurointerv Surg* 2020;12:1045–48; published correction appears in *J Neurointerv Surg* 2021;13:e26 CrossRef
9. Escalard S, Maïer B, Redjem H, et al. **Treatment of acute ischemic stroke due to large vessel occlusion with COVID-19: experience from Paris.** *Stroke* 2020;51:2540–43 CrossRef Medline
10. Pezzini A, Grassi M, Silvestrelli G; STROKCOVID group, et al. **Impact of SARS-CoV-2 on reperfusion therapies for acute ischemic stroke in Lombardy, Italy: the STROKCOVID network.** *J Neurol* 2021; 268:3561–68 CrossRef Medline
11. Qureshi AI, Baskett WI, Huang W, et al. **Utilization and outcomes of acute revascularization treatments in ischemic stroke patients with SARS-CoV-2 infection.** *J Stroke Cerebrovasc Dis* 2022;31:106157 CrossRef Medline

12. Jabbour P, Dmytriw AA, Sweid A, et al. **Characteristics of a COVID-19 cohort with large vessel occlusion: a multicenter international study.** *Neurosurgery* 2022;90:725–33 CrossRef Medline
13. Sawczyńska K, Wrona P, Keósek T, et al. **Mechanical thrombectomy in COVID-19-associated ischaemic stroke: patient characteristics and outcomes in a single-centre study.** *Neurol Neurochir Pol* 2022; 56:163–70 CrossRef Medline
14. The COVID Tracking Project. *Data Download.* 2021
15. Johnston CD, Chen R. **The COVID-19 pandemic and its impact on the southern United States.** *J Comp Fam Stud* 2020;51:314–23 CrossRef
16. Florida Department of Health. Florida Department of Health Open Data. 2020
17. Gupta R, Mouawad NJ, Yi JA. **The impact of the COVID-19 pandemic on vascular surgery: health care systems, economic, and clinical implications.** *Semin Vasc Surg* 2021;34:74–81 CrossRef Medline
18. Rashid Hons M, Gale Hons CP, Curzen Hons N, et al. **Impact of coronavirus disease 2019 pandemic on the incidence and management of out-of-hospital cardiac arrest in patients presenting with acute myocardial infarction in England.** *J Am Heart Assoc* 2020;9:e018379 CrossRef Medline
19. Rednor S, Eisen LA, Cobb JP, et al. **Critical care response during the COVID-19 pandemic.** *Crit Care Clin* 2022;38:623–37 CrossRef Medline
20. Ganesh A, Stang JM, McAlister FA, et al. **Changes in ischemic stroke presentations, management and outcomes during the first year of the COVID-19 pandemic in Alberta: a population study.** *CMAJ* 2022;194:E444–55 CrossRef Medline
21. Legriel S, Badenes R, Engrand N; for NeuroCovid19, et al. **Outcomes in patients with covid-19 with acute encephalopathy and coma: an international prospective study.** *Neurology* 2023;100:e2247–e2258 CrossRef Medline
22. Greene C, Connolly R, Brennan D, et al. **Blood-brain barrier disruption and sustained systemic inflammation in individuals with long COVID-associated cognitive impairment.** *Nat Neurosci* 2024;27:1019 CrossRef Medline
23. Sashindranath M, Nandurkar HH. **Endothelial dysfunction in the brain: setting the stage for stroke and other cerebrovascular complications of COVID-19.** *Stroke* 2021;52:1895–904 CrossRef Medline

# Hydrogel Coils versus Bare Platinum Coils for the Treatment of Ruptured and Unruptured Aneurysms: An Updated Systematic Review and Meta-Analysis of Randomized Controlled Trials

 Jonathan Cortese,  Sherief Ghozy,  Armin Zarrintan, Delal Bektas, Omar M. Al-Janabi, Onam Verma,  Esref Alperen Bayraktar,  Waleed Brinjikji,  Ramanathan Kadirvel, and  David F. Kallmes



## ABSTRACT

**BACKGROUND AND PURPOSE:** Whether hydrogel coils (HGCs) can reduce intracranial aneurysm recurrences when compared with bare platinum coils (BPCs) remains a disputed subject. Thus, we sought to perform a systematic review and meta-analysis to evaluate the efficacy of hydrogel coils in the context of intracranial aneurysm treatment.

**MATERIALS AND METHODS:** Following PRISMA 2020 guidelines, we systematically reviewed PubMed, Scopus, Embase, and Web of Science for randomized controlled trials (RCTs) comparing HGC to BPC. Outcomes of interest were: end of procedure and last follow-up occlusion rates, including complete occlusion and major recurrence, complication rates, morbidity, and mortality. Risk ratios (RRs) and 95% CIs were calculated.

**RESULTS:** The study selection: 5 RCTs, including 2126 patients (HGC  $n = 1064$ , BPC  $n = 1062$ ), were analyzed. HGC showed comparable rates of immediate complete occlusion (RR = 0.89, 95% CI = 0.68–1.16,  $P = .29$ ) and packing attenuation (MD = 27.17, 95% CI = –16.59–70.93,  $P = .12$ ) compared with BPC. At an average 18-month follow-up, HGC significantly reduced major recurrence rates (RR = 0.75, 95% CI = 0.60–0.94,  $P = .03$ ). Complete occlusion rates at the last follow-up were higher for HGC after outlier exclusion (RR = 1.29, 95% CI = 1.18–1.42,  $P < .001$ ). All the outcomes related to complications, including hemorrhagic and thromboembolic complications, were similar between both groups ( $P > .1$  for all). Finally, HGC resulted in similar rates of mRS 0–2 and mortality compared with BPC (RR = 0.98, 95% CI = 0.95–1.01,  $P = .15$  and RR = 0.72, 95% CI = 0.31–1.65,  $P = .33$ , respectively). Only 5 RCTs were included in this meta-analysis, which may limit the generalizability of our findings. The absence of long-term follow-up also limits the assessment of treatment durability.

**CONCLUSIONS:** Our meta-analysis of RCTs suggests that the use of HGC in the endovascular treatment of intracranial aneurysms results in significantly lower rates of recurrence compared with BPC, with both coil types showing similar initial occlusion rates and safety profiles.

**ABBREVIATIONS:** BPC = bare platinum coil; HGC = hydrogel coil; IA = intracranial aneurysm; RCT = randomized controlled trial; RR = risk ratio; RROC = Raymond-Roy Occlusion Classification

Treatment of a ruptured intracranial aneurysm (IA) is a matter of urgency to avoid an early second rupture, which is most often fatal.<sup>1</sup> Treatment of an unruptured aneurysm, on the other hand, is a preventive treatment whose indication is based

on consideration of risk factors for rupture linked to the anatomy of the aneurysm and the patient.<sup>2</sup> Endovascular treatment by coiling was developed in the 1990s and marked an important milestone for patient care. Since the results of the International Subarachnoid Aneurysm Trial, which showed better outcomes with endovascular coiling than with neurosurgical clipping for the treatment of intracranial aneurysms, endovascular coiling remains the preferred treatment for most patients.<sup>3–5</sup>

Minimally invasive aneurysm embolization with coils has been shown to be safe and effective. However, a clear limitation to this technique appeared with aneurysm recurrence occurring in approximately 20%–30% of aneurysms treated with coils.<sup>3,6,7</sup> Aneurysm treatment aims to completely exclude the aneurysm; placement of coils within the aneurysm's sac promotes thrombosis and mitigates the risk of rupture.<sup>8</sup>

Received October 18, 2024; accepted after revision December 30.

From the Departments of Radiology (J.C., S.G., A.Z., E.A.B., W.B., D.F.K.) and Neurologic Surgery (S.G., D.B., R.K.), Mayo Clinic, Rochester, Minnesota; Department of Interventional Neuroradiology—NEURI Brain Vascular Center (J.C.), Bicetre University-Hospital, Le Kremlin-Bicetre, France; Department of Neurology (O.M.A.-J.), Banner-University Medical Center, Tucson, Arizona; and Institute of Medical Education and Research (PGIMER) (O.V.), Chandigarh, India.

Please address correspondence to Jonathan Cortese, MD, Department of Radiology, Mayo Clinic, 200 First St SW, Rochester, MN 55905; e-mail: cortese.jonathan@mayo.edu



Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8697>

Recurrence is the consequence of incomplete and unstable thrombus formation lacking myofibroblast infiltration and collagen deposition in the aneurysm dome and poor endothelialized neointimal layer at the neck of coiled aneurysms.<sup>8</sup>

Because of the high recanalization rates following endovascular coiling and the importance of attenuated packing of the aneurysmal sac in improving long-term occlusion rates, a number of second-generation coils have been developed. The hydrogel coil (HGC; MicroVention) was developed with an expandable polymer that allows improved aneurysm filling.<sup>9</sup> The pressing question of whether HGC offers advantages over the established bare platinum coils (BPCs) in the context of treating intracranial aneurysms remains disputed, though repeatedly evaluated through randomized controlled studies (RCTs).<sup>6,10-12</sup>

A previous meta-analysis of RCTs found that HGC significantly reduced IA recurrences in the midterm follow-up but did not show any significant differences in other outcomes, including complete occlusion at midterm.<sup>13</sup> The authors recommended further exploration of the potential benefits of HGC, particularly second-generation HGC, on complete occlusion and clinical outcomes. Since then, a new RCT with the largest population to date, using second-generation HGC, was published, providing further favorable results for HGC but only in terms of recurrence.<sup>14</sup> In response, we conducted an updated systematic review of the literature and meta-analysis of RCTs to objectively evaluate the effectiveness of HGC compared with BPC in the treatment of IA. This research aims to provide a clear and direct assessment of the angiographic and clinical outcomes associated with these 2 coil types, contributing to a more informed understanding of their respective roles in neuroendovascular interventions.

## MATERIALS AND METHODS

### Search Strategy

On January 17, 2024, following PRISMA recommendations for systematic reviews,<sup>15</sup> we conducted a comprehensive literature search within the Nested Knowledge Autolit software based on our prespecified protocol. The search covered 4 databases from their inception: PubMed, Embase, Web of Science, and Scopus. Only English-language studies were included.

For each database, we employed different combinations of keywords and/or MeSH terms. The terms included “Hydrogel,” “Hydrocoil,” “HydroSoft,” “HydroFrame,” “HydroFill,” “coil,” “aneurysm,” and “randomized.” Boolean operators (AND, OR) were used to ensure a comprehensive retrieval of studies related to hydrogel-coated coils in the treatment of aneurysms.

In addition, we conducted a manual search through the references of included articles to identify any potentially missed relevant studies. All search results were exported to a Nested Knowledge Autolit for deduplication and screening according to the inclusion criteria.

### Screening Process

We included all original studies fulfilling our predetermined Population, Intervention, Comparison, and Outcome (PICO). The population was patients with IA, intervention treatment with HGC, and control group treatment with BPC; the outcomes of interest were immediate and last follow-up complete occlusion

and adequate occlusion, major recurrence, retreatment rate, favorable outcome (mRS = 0–2), mortality rates, and complication rates (including ischemic and hemorrhagic complications). RCTs comparing treatment outcomes for IA by using BPC versus HGC were included without restrictions on publication date, country of origin, aneurysm location, or rupture status. All other study designs were excluded. Two authors (S.G. and O.M.A.-J.) did the title and abstract screening against the predefined criteria. This was followed by a full-text screening of any retained studies of the first screening step. In both stages, the senior author was consulted to resolve any conflicts in the decisions. Complete occlusion and adequate occlusion were defined according to specific criteria. For complete occlusion, Raymond-Roy Occlusion Classification (RROC) grade I was used as a benchmark. Additionally, for adequate occlusion, defined as complete aneurysm occlusion along with a neck remnant, RROC grades I and II were considered.<sup>16</sup>

### Data Extraction

The extraction of selected articles was performed by 2 authors (J.C. and A.Z.). The extracted data included study characteristics, baseline data of the included patients, and the aforementioned outcomes of interest. The data that support the findings of this study are available from the corresponding author upon reasonable request.

### Risk of Bias

The “Cochrane RoB 2: a revised tool for assessing risk of bias in randomized trials” was used to assess the risk of bias, with 2 independent reviewers evaluating all studies (A.Z. and O.V.).<sup>17</sup> Two authors evaluated the quality of each study, and it was adjudicated by a third one when needed.

### Statistical Analysis

In this study, we performed a random-effects meta-analysis by using R software Version 4.3.1 and the “meta” statistical package.<sup>18</sup> We calculated risk ratios (RRs) along with their respective 95% CIs. The confidence intervals for the random-effects estimates were derived by using the Restricted Maximum Likelihood estimator, incorporating Jackson’s modification of the Hartung-Knapp/Sidik-Jonkman variance correction, and the “Hybrid method 2” in Jackson et al<sup>19</sup> as an ad hoc correction. To handle studies with zero events in both treatment groups (double zeros), we applied the treatment arm continuity correction method not to discard the information in such cases.<sup>20,21</sup>

Heterogeneity was assessed by using the Cochran Q test and the  $I^2$  statistic. A *P* value < .05 for Cochran Q was considered indicative of statistically significant heterogeneity, while  $I^2$  values >50% were interpreted as substantial heterogeneity. To explore sources of heterogeneity, if substantial, we conducted sensitivity analysis by using the leave-one-out method, examining the impact of excluding each study on the overall effect size. Because of the small number of studies (fewer than 10) included in the meta-analysis, we were unable to perform Egger regression to assess publication bias or conduct meta-regression.<sup>22</sup>

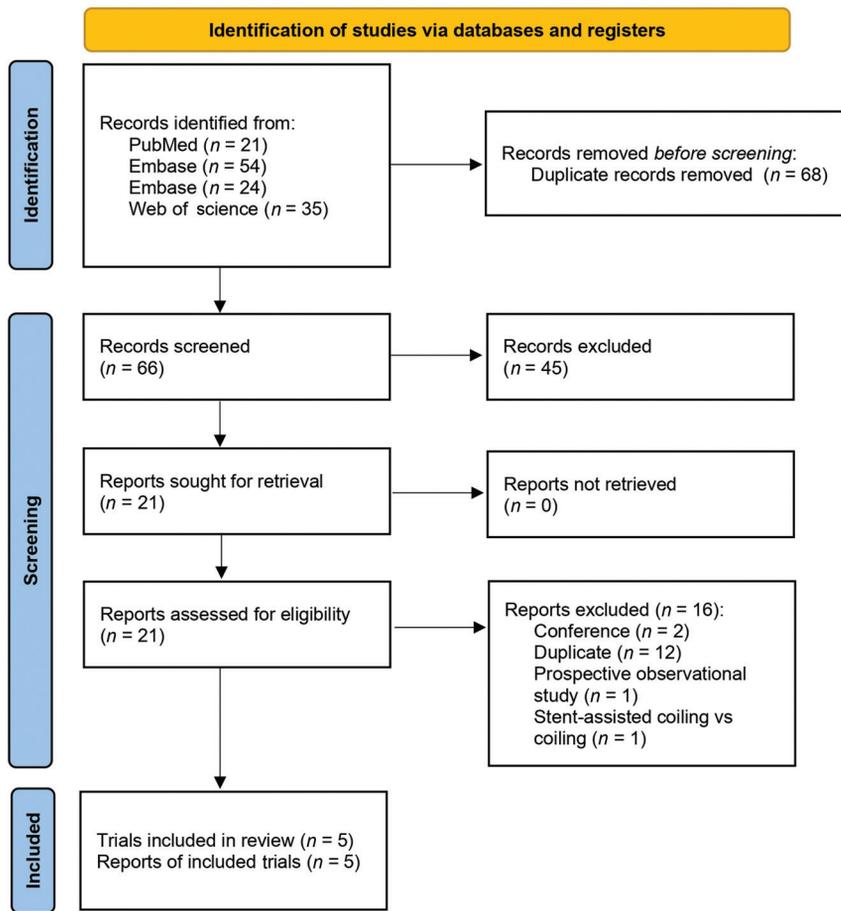


FIG 1. PRISMA flowchart of the included studies.

Study	Risk of bias domains					Overall
	D1	D2	D3	D4	D5	
White et al, 2011	+	+	+	+	+	+
Poncyłjusz et al, 2014	-	+	-	-	+	×
Raymond et al, 2015	+	+	+	+	+	+
Taschner et al, 2018	+	-	+	+	+	-
Bendok et al, 2020	+	+	+	+	+	+

Domains:  
D1: Bias arising from the randomization process.  
D2: Bias due to deviations from intended intervention.  
D3: Bias due to missing outcome data.  
D4: Bias in measurement of the outcome.  
D5: Bias in selection of the reported result.

Judgement  
× High  
- Some concerns  
+ Low

FIG 2. Results of the methodologic quality assessment of the included studies based on Cochrane RoB 2. "Traffic light" plots of the domain-level judgments for each individual result.

## RESULTS

### Search and Screening Results

A total of 134 records were identified across 4 databases: PubMed ( $n = 21$ ), Embase ( $n = 54$ ), Scopus ( $n = 24$ ), and Web of Science ( $n = 35$ ). After 68 duplicate records were removed, 66 records remained for screening. During the screening of titles and abstracts, 45 records were excluded for various reasons,

including a comparison between clipping versus coiling, secondary analyses, book chapters, letters, and studies unrelated to the topic.

From the remaining records, 21 full-text reports were sought for retrieval, all of which were successfully retrieved and assessed for eligibility. After further evaluation, 16 reports were excluded for reasons including conference abstracts ( $n = 2$ ), duplicates ( $n = 12$ ), prospective observational studies ( $n = 1$ ), and comparisons of stent-assisted coiling versus coiling ( $n = 1$ ). Ultimately, 5 studies were included in the final review (Fig 1).

### Study Characteristics and Risk of Bias

All 5 studies included in our analysis were RCTs. The size of the included studies ranged from 96 patients to 600 patients, with a total of 2126 patients enrolled and pooled for the analysis (hydrogel  $n = 1064$ , BPC  $n = 1062$ ). Study and patient characteristics are detailed in Supplemental Data.

RCTs included in our analysis were deemed to have a "low" risk of bias for 3 of them, "some concerns" for 1 of them, and the last study was considered to have a "high" risk (Fig 2).

### End of Procedure Occlusion Rates

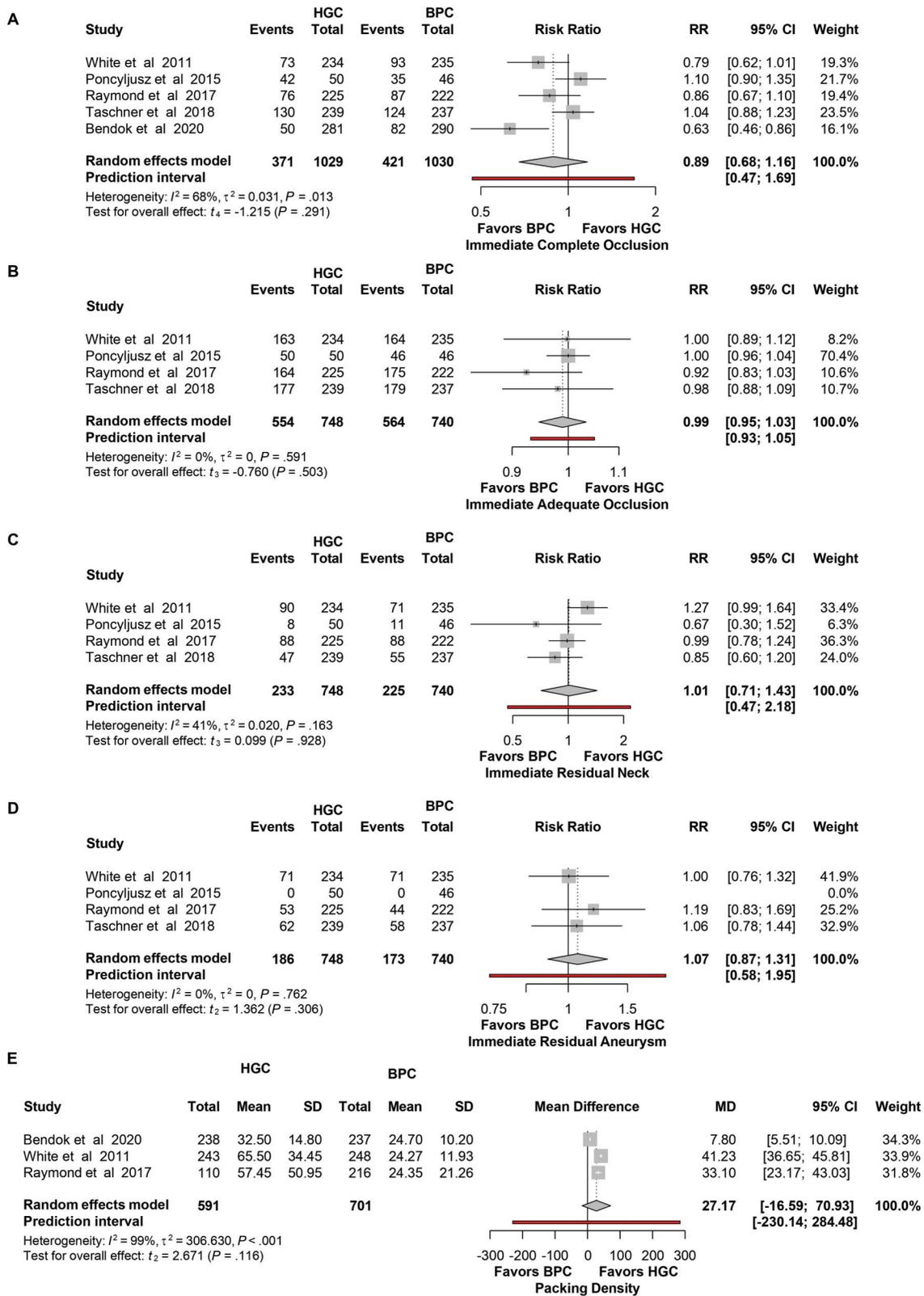
All 5 RCTs, with 2059 patients, reported the rate of immediate complete occlusion. HGC resulted in comparable rates of immediate complete occlusion compared with BPC (RR = 0.89, 95% CI = 0.68–1.16,  $P = .29$ ). There was substantial heterogeneity among the included studies ( $I^2 = 68\%$ ,  $P$  value = .01) (Fig 3A). The sensitivity analysis with the leave-one-out method resulted in similar results, regardless of what the study omitted (Supplementary Data).

HGC also resulted in comparable rates of immediate adequate occlusion compared with BPC (RR = 0.99, 95% CI = 0.95–1.03,  $P$  value = .50), which

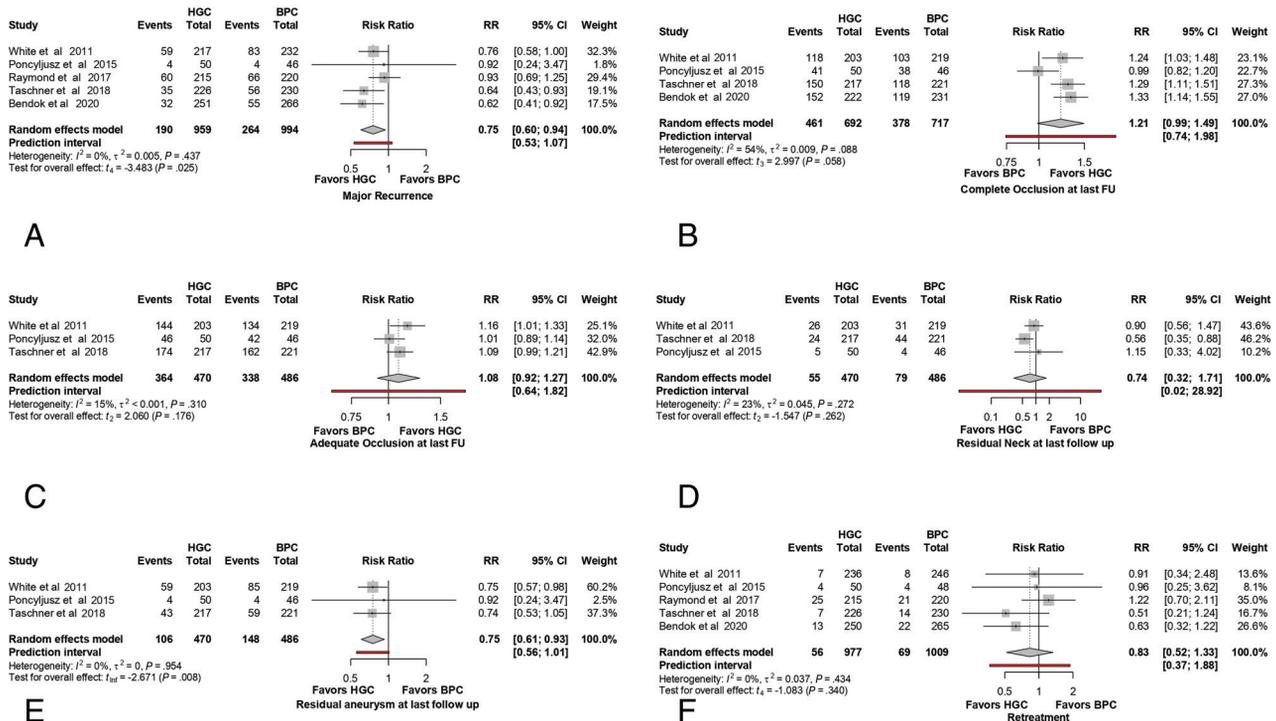
was reported by 4 RCTs, with 1488 patients. There was no heterogeneity among the included studies ( $I^2 = 0\%$ ,  $P = .59$ ) (Fig 3B).

Furthermore, immediate residual neck and immediate residual aneurysm were comparable between both groups (Fig 3C, -D).

Finally, 3 RCTs, with 1291 patients, reported the packing attenuation at the end of the procedure. HGC resulted in comparable rates of packing attenuation compared with BPC (MD =



**FIG 3.** Forest plot of the random effects proportion meta-analysis of the reported rates of the end-of-procedure occlusion outcomes. A, Immediate complete occlusion. B, Immediate adequate occlusion. C, Immediate residual neck. D, Immediate residual aneurysm. E, Packing attenuation.



**FIG 4.** Forest plot of the random effects proportion meta-analysis of the reported rates of the last follow-up occlusion outcomes. A, Major recurrence. B, Complete occlusion. C, Adequate occlusion. D, Residual neck. E, Residual aneurysm. F, Retreatment.

27.17, 95% CI =  $-16.59-70.93$ ,  $P = .12$ ). There was substantial heterogeneity among the included studies ( $I^2 = 99\%$ ,  $P < .001$ ) (Fig 3E).

### Last Follow-Up Occlusion Rates

Average last follow-up in all studies was 18 months (range 12–24 months).

All 5 RCTs, with 1953 patients, reported the rate of major recurrence at the last follow-up. HGC resulted in significantly reduced rates of major recurrence at the last follow-up compared with BPC (RR = 0.75, 95% CI = 0.60–0.94;  $P = .03$ ). There was no heterogeneity among the included studies ( $I^2 = 0\%$ ,  $P$  value = .44) (Fig 4A). It should be noted that slightly different definitions of major recurrence were used. The Hydrogel Endovascular Aneurysm Treatment (HEAT) trial defined it as an increase on the RROC scale from 1 to 3, 2 to 3, or as a grade 3 that showed progression on the Meyer scale. For the GREAT trial and Poncyjusz et al,<sup>10</sup> major recurrence was defined as any change from complete aneurysm occlusion or neck remnant at the end of the index procedure to residual aneurysm at angiographic follow-up. The HELP and Patients Prone to Recurrence After Endovascular Treatment (PRET) trials defined it as a recurrence sufficiently large enough to technically allow placement of further coils.

Four RCTs, with 1409 patients, reported the rate of complete occlusion at the last follow-up. HGC resulted in similar rates of complete occlusion at the last follow-up compared with BPC (RR = 1.21, 95% CI = 0.99–1.48,  $P = .06$ ). There was substantial heterogeneity among the included studies ( $I^2 = 54\%$ ,  $P = .09$ ) (Fig 4B). When conducting the sensitivity analysis, the omission of the outlier study led to a significantly higher rate of complete

occlusion at the last follow-up for the HGC group compared with BPC (RR = 1.29, 95% CI = 1.18–1.42,  $P < .001$ ) (Supplementary Data).

Adequate occlusion at the last follow-up was similar between both groups (RR = 1.08, 95% CI = 0.92–1.27,  $P = .31$ ), which was reported in 3 RCTs, with 956 patients. There was no significant heterogeneity among the included studies ( $I^2 = 15\%$ ,  $P = .310$ ) (Fig 4C). Similarly, residual neck at the last follow-up was similar between both groups (RR = 0.74, 95% CI = 0.32–1.71,  $P = .26$ ) (Fig 4D).

Three RCTs, with 956 patients, also reported the rate of residual aneurysm at the last follow-up. HGC resulted in significantly reduced residual aneurysm at the last follow-up rates compared with BPC (RR = 0.75, 95% CI = 0.61–0.93,  $P = .01$ ). There was no heterogeneity among the included studies ( $I^2 = 0\%$ ,  $P = .95$ ) (Fig 4E).

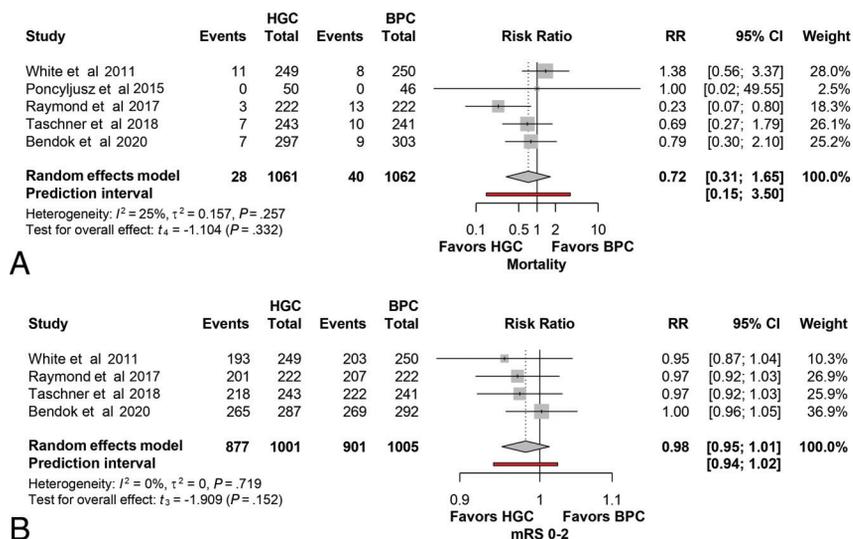
Finally, 5 RCTs, with 1986 patients, reported the rate of retreatment. HGC resulted in similar rates of retreatment compared with BPC (RR = 0.83, 95% CI = 0.52–1.33,  $P = .34$ ). There was no heterogeneity among the included studies ( $I^2 = 0\%$ ,  $P = .43$ ) (Fig 4F).

### Complications

All the outcomes related to complications—hemorrhagic complications, thromboembolic complications, all strokes, coils migration, hydrocephalus, procedural complications, total complications—were similar between both groups (Supplemental Data). There was no heterogeneity among the included studies for each outcome.

### Clinical Outcomes

All 5 RCTs, with 2123 patients, reported the rate of mortality. HGC resulted in similar rates of mortality compared with BPC



**FIG 5.** Forest plot of the random effects proportion meta-analysis of the reported rates of mortality (A) and morbidity (mRS 0–2; B).

(RR = 0.72, 95% CI = 0.31–1.65,  $P = .33$ ). There was no significant heterogeneity among the included studies ( $I^2 = 25\%$ ,  $P = .26$ ) (Fig 5A).

Four RCTs, with 2006 patients, reported the rate of mRS 0–2. HGC resulted in similar rates of mRS 0–2 compared with BPC (RR = 0.98, 95% CI = 0.95–1.01,  $P = .15$ ). There was no heterogeneity among the included studies ( $I^2 = 0\%$ ,  $P$  value = .72) (Fig 5B).

## DISCUSSION

Our meta-analysis, including 5 RCTs comparing HGC to BPC, demonstrates that HGC provides a similar safety profile with superior long-term durability, evidenced by reduced recurrence rates. These results are clinically significant, particularly for unruptured aneurysms or those prone to recurrence, where retreatment rates can pose additional risks to patients.

### Key Findings and Clinical Implications

The primary finding of our meta-analysis is the significantly lower rate of major recurrence at midterm follow-up with HGC compared with BPC. This observation is consistent with previous studies, such as the GREAT and HEAT trials, which reported reductions in recurrence rates and trends toward decreased retreatment for HGCs.<sup>11,14</sup> Moreover, a systematic review of first- and second-generation HGCs compared with BPCs highlighted the reduced recurrence and favorable long-term outcomes, supporting their use in clinical practice.<sup>23</sup> The reduction in recurrence with HGCs may be attributed to the unique hydrogel coating, which expands upon contact with blood, resulting in increased packing attenuation and a more stable thrombus formation within the aneurysm sac.<sup>8</sup>

Other systematic reviews and meta-analyses have reached similar conclusions regarding recurrence rates.<sup>13,24,25</sup> Nonetheless, we conducted an up-to-date analysis to specifically address the comparison between HGC and BPC in RCTs only. This approach allowed us to aggregate the largest randomized population,

yielding more reliable results on additional occlusion outcomes and clinical outcomes. We found a significantly higher rate of complete occlusion at midterm follow-up with HGC compared with BPC, a result not achieved in previous meta-analyses.<sup>13</sup> Earlier studies included either retrospective analyses<sup>23</sup> or varied types of bioactive coils.<sup>24,25</sup>

Our analysis also found no significant differences in initial complete occlusion rates between HGC and BPC, which suggests that HGCs do not compromise immediate procedural efficacy. While this similarity in occlusion rates might initially seem to negate the benefits of HGC, the reduced recurrence rates highlight the importance of long-term stability over immediate outcomes. This could be particularly valuable in aneurysms with a higher baseline risk

of recanalization or in cases where complete initial occlusion is challenging to achieve.<sup>12</sup>

### Safety Profile of Hydrogel Coils

The safety profile of HGCs was comparable to BPCs across various adverse events, including hemorrhagic complications (RR = 0.73), coil migration (RR = 1.53), hydrocephalus (RR = 1.30), and thromboembolic complications (RR = 0.73), with no significant heterogeneity among the studies for these outcomes (all  $I^2 = 0\%$ ). This is consistent with findings from a meta-analysis of randomized controlled trials, which reported similar complication rates between hydrogel and bare platinum coils across a pooled sample of 1526 patients.<sup>13</sup> Furthermore, the GREAT trial also reported comparable overall procedural complication rates and similar adverse composite outcomes between HGC and BPC groups, demonstrating that HGCs do not compromise safety even in high-risk populations.<sup>11</sup> The PRET trial, which focused on large and recurrent aneurysms, found similar safety outcomes between the 2 coil types, highlighting that HGCs are a safe alternative even for challenging aneurysm profiles.<sup>12,26</sup> These findings indicate that the benefits of HGCs in reducing recurrence are not achieved at the cost of increased procedural risks. The stability of these results, even in high-risk populations and across different centers, reinforces the safety of HGCs as an alternative to BPCs for intracranial aneurysm treatment.

### Histologic Considerations and Mechanistic Insights

The reduced recurrence rates observed with HGCs can be attributed to the distinct biologic properties of the hydrogel coating. Upon contact with blood, the hydrogel expands, leading to better aneurysm sac filling and a denser, more stable thrombus that reduces porosity and minimizes the risk of coil compaction.<sup>8</sup> This structural stability provides an optimal scaffold for cellular infiltration, particularly by myofibroblasts, which play a critical role in stabilizing the aneurysm through collagen synthesis and fibrous tissue deposition.<sup>27</sup> The increased collagen deposition at

the aneurysm neck strengthens the neointima layer, effectively sealing the aneurysm and reducing the likelihood of delayed recanalization. This suggests that HGCs not only offer mechanical stability but also promote a more organized healing process at the aneurysm neck. Experimental studies have confirmed these histologic changes, suggesting that the unique properties of HGCs may provide enhanced long-term stability compared with BPCs.<sup>8</sup>

### Limitations and Future Directions

Several limitations should be noted. First, the inclusion of only 5 RCTs limits the generalizability of our findings. While certain outcomes, such as major recurrence and residual aneurysm rates, showed no heterogeneity ( $I^2 = 0\%$ ), others, such as immediate and long-term complete occlusion, displayed moderate heterogeneity ( $I^2 = 68\%$  and  $54\%$ , respectively). High heterogeneity was observed for packing attenuation at the end of the procedure ( $I^2 = 99\%$ ), indicating considerable variability across studies, likely due to differences in procedural techniques and aneurysm characteristics. This variability may limit the robustness of our conclusions for some end points. Additionally, incomplete reporting of long-term outcomes in several studies could limit the assessment of treatment durability. Future studies should focus on aneurysm subtypes in different clinical scenarios, such as ruptured, unruptured, and reanalyzed aneurysms, to provide more comprehensive insights.

### CONCLUSIONS

Our meta-analysis supports the use of HGC as a safe and effective alternative to BPC for the treatment of intracranial aneurysms, particularly in reducing major recurrence rates without compromising safety. Given the comparable complication profiles and potential for improved midterm outcomes, HGCs should be considered in clinical scenarios where long-term aneurysm stability is critical, such as wide-neck aneurysms or cases with a high risk of recurrence. Further studies are warranted to confirm these findings and explore their role in different aneurysm types and clinical contexts.

### ACKNOWLEDGMENTS

J.C. received educational grants from the French Society of Neuroradiology (SFNR), the French Society of Radiology, the Philippe Foundation, and the Foundation Thérèse and René Planiol.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

### REFERENCES

- Vega C, Kwoon JV, Lavine SD. **Intracranial aneurysms: current evidence and clinical practice.** *Am Fam Physician* 2002;66:601–08 Medline
- Wiebers DO. **Unruptured intracranial aneurysms: natural history, clinical outcome, and risks of surgical and endovascular treatment.** *The Lancet* 2003;362:103–10 CrossRef
- Molyneux A, Kerr R, Stratton I; International Subarachnoid Aneurysm Trial (ISAT) Collaborative Group, et al. **International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised trial.** *Lancet* 2002;360:1267–74 CrossRef Medline
- Molyneux AJ, Birks J, Clarke A, et al. **The durability of endovascular coiling versus neurosurgical clipping of ruptured cerebral aneurysms: 18 year follow-up of the UK cohort of the International Subarachnoid Aneurysm Trial (ISAT).** *Lancet* 2015;385:691–97 CrossRef Medline
- Brinjikji W, Rabinstein AA, Nasr DM, et al. **Better outcomes with treatment by coiling relative to clipping of unruptured intracranial aneurysms in the United States, 2001–2008.** *AJNR Am J Neuroradiol* 2011;32:1071–75 CrossRef Medline
- White PM, Lewis SC, Gholkar A; HELPS Trial Collaborators, et al. **Hydrogel-coated coils versus bare platinum coils for the endovascular treatment of intracranial aneurysms (HELPS): a randomised controlled trial.** *Lancet* 2011;377:1655–62 CrossRef Medline
- Ferns SP, Sprengers MES, van Rooij WJ, et al. **Coiling of intracranial aneurysms: a systematic review on initial occlusion and reopening and retreatment rates.** *Stroke* 2009;40:e523–29 CrossRef Medline
- Brinjikji W, Kallmes DF, Kadirvel R. **Mechanisms of healing in coiled intracranial aneurysms: a review of the literature.** *AJNR Am J Neuroradiol* 2015;36:1216–22 CrossRef Medline
- Kallmes DF, Fujiwara NH. **New expandable hydrogel-platinum coil hybrid device for aneurysm embolization.** *AJNR Am J Neuroradiol* 2002;23:1580–88 Medline
- Poncyjusz W, Zarzycki A, Zwarzany Ł, et al. **Bare platinum coils vs. HydroCoil in the treatment of unruptured intracranial aneurysms—a single center randomized controlled study.** *Eur J Radiology* 2015;84:261–65 CrossRef Medline
- Taschner CA, Chapot R, Costalat V, et al. **Second-generation hydrogel coils for the endovascular treatment of intracranial aneurysms: a randomized controlled trial.** *Stroke* 2018;49:667–74 CrossRef Medline
- Raymond J, Klink R, Chagnon M, et al. **Hydrogel versus bare platinum coils in patients with large or recurrent aneurysms prone to recurrence after endovascular treatment: a randomized controlled trial.** *AJNR Am J Neuroradiol* 2017;38:432–41 CrossRef Medline
- Xue T, Chen Z, Lin W, et al. **Hydrogel coils versus bare platinum coils for the endovascular treatment of intracranial aneurysms: a meta-analysis of randomized controlled trials.** *BMC Neurol* 2018;18:167 CrossRef Medline
- Bendok BR, Abi-Aad KR, Ward JD; HEAT Study Investigators, et al. **The Hydrogel Endovascular Aneurysm Treatment Trial (HEAT): a randomized controlled trial of the second-generation hydrogel coil.** *Neurosurgery* 2020;86:615–24 CrossRef Medline
- Moher D, Liberati A, Tetzlaff J; PRISMA Group, et al. **Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.** *Open Med* 2009;3:e123–30 CrossRef Medline
- Mascitelli JR, Moyle H, Oermann EK, et al. **An update to the Raymond-Roy Occlusion Classification of intracranial aneurysms treated with coil embolization.** *J Neurointerv Surg* 2015;7:496–502 CrossRef Medline
- Sterne JAC, Savović J, Page MJ, et al. **RoB 2: a revised tool for assessing risk of bias in randomised trials.** *BMJ* 2019;366:l4898 CrossRef Medline
- Schwarzer G. **Meta: An R package for meta-analysis.** *R News* 2007;7:40–45
- Siemens W, Meerpohl JJ, Rohe MS, et al. **Reevaluation of statistically significant meta-analyses in advanced cancer patients using the Hartung-Knapp method and prediction intervals—a methodological study.** *Res Synth Methods* 2022;13:330–41 CrossRef Medline
- Wei J-J, Lin E-X, Shi J-D, et al. **Meta-analysis with zero-event studies: a comparative study with application to COVID-19 data.** *Mil Med Res* 2021;8:41 CrossRef Medline
- Ren Y, Lin L, Lian Q, et al. **Real-world performance of meta-analysis methods for double-zero-event studies with dichotomous outcomes using the Cochrane database of systematic reviews.** *J Gen Intern Med* 2019;34:960–68 CrossRef Medline
- Chandler J, Cumpston M, Li T, et al. *Cochrane Handbook for Systematic Reviews of Interventions.* Wiley; 2019

23. Abi-Aad KR, Rahme RJ, Patra DP, et al. **Clinical outcomes of first- and second-generation hydrogel coils compared with bare platinum coils: a systematic literature review.** *Neurosurg Rev* 2022; 45:1873–82 CrossRef Medline
24. Broeders JA, Ahmed Ali U, Molyneux AJ, et al. **Bioactive versus bare platinum coils for the endovascular treatment of intracranial aneurysms: systematic review and meta-analysis of randomized clinical trials.** *J Neurointerv Surg* 2016;8:898–908 CrossRef Medline
25. Zhang J, Jiang G, Song Z, et al. **Efficacy and safety of different bioactive coils in intracranial aneurysm interventional treatment, a systematic review and Bayesian network meta-analysis.** *Brain Sci* 2022;12:1062 CrossRef Medline
26. Raymond J, Roy D, White PM; Icone Collaborative Group, et al. **A randomized trial comparing platinum and hydrogel-coated coils in patients prone to recurrence after endovascular treatment (the PRET Trial).** *Interv Neuroradiol* 2008;14:73–83 CrossRef Medline
27. D’Urso M, Kurniawan NA. **Mechanical and physical regulation of fibroblast-myofibroblast transition: from cellular mechanoreponse to tissue pathology.** *Front Bioeng Biotechnol* 2020;8:609653 CrossRef Medline

# Endovascular Treatment of Wide-Neck Intracranial Aneurysms Using the Novel Contour Neurovascular System: 5-Year Follow-Up

Howra Ktayan,  Christopher Yusuf Akhunbay-Fudge,  Atul Kumar Tyagi, Kenan Deniz, and Tufail Patankar



## ABSTRACT

**BACKGROUND AND PURPOSE:** The Contour neurovascular embolization device is a novel way to treat wide-neck bifurcation aneurysms (WNBA), which often pose considerable treatment challenges. In this study, we aim to evaluate the efficacy and safety profile of this device.

**MATERIALS AND METHODS:** Prospective clinical and radiologic data were collected for all patients treated with the Contour device at our center, between January 2017 and December 2018. All patients were treated electively, and aneurysms were unruptured.

**RESULTS:** Fourteen patients were recruited, and the device was successfully deployed in 11 patients. All patients were women with a mean age of 65 years. Four basilar tip, 2 internal carotid, 3 middle cerebral, 1 anterior communicating, and 1 superior cerebellar artery aneurysms were treated. The mean aneurysmal size was 6 mm (width) × 7.6 mm (height), with 4.1 mm neck. Follow-up imaging included DSA, MRA, and CTA. For the 9 patients available at year 2 follow-up, 5 showed improved occlusion class over time with 8 of 9 having adequate occlusion defined by class 1 and 2 of the Raymond-Roy (R-R) classification system. Eight patients were available for 3-year follow-up: 7 patients had stable occlusion class including 3 patients with stable complete occlusion (R-R class 1). One patient had worsening of R-R occlusion class from 1 to 2 and subsequently presented with acute subarachnoid hemorrhage. Follow-up data 5 years after implantation were available for 8 patients (including the ruptured and re-treated patient): 7 patients had adequate occlusion (R-R class 1 and 2), however, 1 patient had worsened from R-R class 1 to R-R class 2. Four patients had complete occlusion.

**CONCLUSIONS:** Results demonstrate progressive occlusion of wide-neck aneurysms over the first 2 years, but we have also demonstrated worsening of R-R occlusion class in some aneurysms that were previously completely occluded. Our results suggest that the Contour device is a good option in WNBA, however, it appears that patients must be followed up for a minimum of 5 years.

**ABBREVIATIONS:** CNS = Contour Neurovascular System; GCS = Glasgow Coma Scale; R-R = Raymond-Roy; WEB = Woven EndoBridge; WNBA = wide-neck bifurcation aneurysms

Endovascular treatment of intracranial aneurysms has become the mainstay of treatment in most neurosurgical centers, particularly in the case of acutely ruptured aneurysms. In terms of treating wide-neck bifurcation aneurysms (WNBA), traditional coiling methods are suboptimal, hence the development of intrasaccular devices. In 2019, a multicenter US study looked at wide-neck MCA and basilar tip aneurysms treated by using endovascular techniques such as simple, balloon, or stent-assisted coiling. They demonstrated adequate occlusion rates of between

40% and 63% with retreatment rates of 8.7%.<sup>1</sup> Similar findings have repeatedly been published, demonstrating the need for novel strategies in the treatment of WNBA.<sup>2,3</sup> New devices have been developed, such as Woven EndoBridge (WEB; Microvention) and LUNA devices, with better occlusion rates than traditional coiling techniques.<sup>4-6</sup>

In 2020, we published the first human series on treating WNBA with the Contour Neurovascular System (CNS; Cerus Endovascular), and demonstrated a complete occlusion rate of 56% at 1 year.<sup>7</sup> Although we demonstrated satisfactory rates of occlusion at 1 year, it is essential that patients are followed up for longer, to establish the true efficacy, reliability, and safety of this novel device. In this paper, we share our results after 5-year follow-up on 8 of the first 11 patients who had the CNS device implanted at our center. To our knowledge, this is the first study to publish up to 5-year follow-up on patients treated with the Contour device.

Received November 25, 2023; accepted after revision December 26, 2024.

From the Departments of Neurosurgery (H.K., C.Y.A.-F., A.K.T., K.D.), and Interventional Neuroradiology (T.P.), Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom.

Please address correspondence to Howra Ktayan, Neurosurgery Department, G Floor, Jubilee Wing, Leeds General Infirmary, Great George St, LS1 3EX; e-mail: h.ktayan@nhs.net



Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8641>

## SUMMARY

**PREVIOUS LITERATURE:** Several studies have evaluated the use of the Contour device in wide-neck bifurcation aneurysms. In 2021, the CERUS study reported adequate occlusion rates of 84%. In 2022, the largest series to date, which included 60 aneurysms treated with the Contour device, demonstrated adequate occlusion rates of 89% at 1 year. These results are comparable with those of the Woven EndoBridge device, which has demonstrated adequate occlusion rates between 73% and 89%. Current literature suggests that the Contour device is a good option in treating wide-neck aneurysms, however, there is limited follow-up on these patients.

**KEY FINDINGS:** We demonstrate an adequate occlusion rate of 87.5% at 5 years. However, our series includes 2 patients with worsening rates of occlusion, 1 of whom presented with acute rupture and SAH 3 years after treatment. We have seen a worsening occlusion rate of 18% at 5 years.

**KNOWLEDGE ADVANCEMENT:** We have highlighted the importance of following patients for at least 5 years, regardless of occlusion status. Furthermore, follow-up imaging should be sensitive for regrowing aneurysms (catheter and CTA), rather than MRI that is susceptible to marker artifact at the neck.

## MATERIALS AND METHODS

### Case Selection

As described in our first in human series with the CNS,<sup>7</sup> before treatment, all cases were discussed in the neurovascular multidisciplinary team meeting. We identified all patients with aneurysms suitable for endovascular treatment, those that had a wide neck, and were unruptured. We defined wide-neck aneurysms as having a neck diameter of  $\geq 4$  mm or a dome-to-neck ratio of  $< 2$ .<sup>2,8</sup> All treatment options were presented to the patients, and those consenting to the Contour device were treated with it and enrolled in the study. We collected prospective clinical and radiologic data for all patients treated with the CNS between January 2017 and December 2018. All cases were treated on an elective basis and informed consent was obtained along with local ethics board approval.

### Analysis

Baseline clinical data including patient demographics, baseline Glasgow Coma Scale (GCS), and mRS were collected. Follow-up mRS at 6 weeks and any complications were recorded, along with immediate posttreatment, 6-month, 1-year, and 2-year radiologic follow-up (in the form of conventional angiography); for detailed information, refer to our prior publication.<sup>7</sup> For the present study, 3-year, 4-year, and 5-year radiologic follow-up (in the form of conventional angiography, contrast-enhanced MRA, or CTA performed in the arterial phase) was recorded. Occlusion

was quantified by using the recognized Raymond-Roy (R-R) class classification.<sup>9</sup> Class 1 denotes complete occlusion, class 2 denotes neck remnant, and class 3 denotes residual aneurysm. We considered R-R class 1 and 2 as satisfactory outcomes, given the good clinical outcomes described associated with these classes in the literature.<sup>10</sup> Device migration was reviewed on postprocedural imaging and device configuration was reviewed by using conventional angiography or CTA.

### Brief Procedural Technique

All cases were performed by the senior author under general anesthetic via standard common femoral artery puncture and an 8F sheath insertion. Patients did not receive any antiplatelet therapy before the procedure, and we did not perform platelet testing. Intraoperatively, 500 mg of IV aspirin was given and patients were discharged with 6 weeks of 75 mg aspirin once per day. For detailed procedural steps, please refer to our prior publication.<sup>7</sup>

## RESULTS

### Cohort Summary

We initially identified 14 patients suitable for treatment; however, we were only able to place the device in 11 of those. The 3 failed cases are discussed in more detail in our previous publication<sup>7</sup> and a summary is provided in Table 1.

**Table 1: Summary of failed cases with the CNS as previously published by Akhunbay-Fudge et al<sup>7</sup>**

Aneurysm Location	Aneurysm Dimensions			AR	DNR	Procedural Difficulties
	Neck (mm)	Width (mm)	Height (mm)			
MCA	3.3	4.6	4.4	1.3	1.4	Unable to catheterize with XT 27 catheter so coils were attempted. Despite eventual catheterization with Echelon 10 microcatheter, balloon inflation, and attempts with multiple coils, the procedure was thought to be high risk of stroke and abandoned
Pericallosal	3.7	6.6	5.5	1.5	1.8	Inability to catheterize aneurysm with XT 27 catheter, therefore successfully reverted to coiling
MCA	5.1	6.3	7.2	1.4	1.2	An XT 27 was navigated into the aneurysm. There were 2 failed trials with 2 11-mm Contour devices. Subsequently a 9 × 4 mm WEB was used but was not suitable and an 8 × 3 mm WEB was used instead with good clinical and radiologic outcomes

**Note:**—AR indicates aspect ratio (height/neck); DNR, dome/neck ratio (width/neck).

Our series therefore includes 11 patients, treated with the CNS device between February 2017 and February 2018. They were all women with a mean (standard deviation) age of 65.0 (6.4) years. All patients were treated on an elective basis for incidentally found aneurysms with 1 patient (patient 1) having previously experienced subarachnoid hemorrhage from a different aneurysm that had been coiled. All patients had a pretreatment GCS of 15 and an admission mRS of 0. In total, 4 basilar tip, 2 ICA, 3 MCA, 1 anterior communicating artery, and 1 superior cerebellar artery aneurysms were treated (Supplemental Data). The mean aneurysmal size was 6 mm (width) × 7.6 mm (height), with 4.1 mm neck. In all 11 cases only 1 CNS device was used, and no adjuvant devices were required at the time of treatment. After 6 months, 2 patients left the study: 1 patient did not want further follow-up and the other patient developed frontal lobe dementia. After 3 years, 1 patient died due to medical deterioration unrelated to the aneurysm. Three years after the procedure, 1 patient presented with acute SAH and underwent stent-assisted coiling of the previously treated aneurysm. One patient did not have an appropriately sized Contour resulting in persistent filling of the aneurysm for 4 years. All other aneurysms remained stable at 4- and 5-year follow-up, with no migration of the device.

### Clinical Outcomes and Complications

As described in our prior publication,<sup>7</sup> of the 15 patients, there were 2 cases of thromboembolic events (14%), however, there were no cases of postoperative ipsilateral major stroke, aneurysmal rupture, or death due to neurologic causes. Patient 6 (basilar tip aneurysm) presented 1 week after the procedure with TIA symptoms and on MRI was found to have 2 small infarcts in the right cerebellar and right occipital lobes. All symptoms resolved completely, the patient had no residual neurologic deficits (NIHSS 0 and mRS 0), and was discharged with 6 weeks of aspirin and 3 months of clopidogrel. Patient 10 had 2 failed attempts at CNS placement before the final 11-mm device was suitably sited. Soon after waking, the patient was found to be hemiplegic. Immediate repeat angiography demonstrated an MCA thrombus, which was treated with 5 mg IV abciximab. After this, the patient had no deficits (NIHSS 0 and mRS 0) and was discharged with 2 months of aspirin and clopidogrel. The thrombus was not related to the detachment zone but had formed on the device during the procedure. The operator had failed to recognize the thrombus, and initially thought it was the device itself.

### Radiologic Outcomes (Supplemental Data)

**Immediately Postprocedure.** Immediate postoperative angiography demonstrated complete occlusion, class 1 as defined by

the R-R classification was seen in 1 patient (9%), a small neck remnant (R-R class 2) in 2 patients (18%), and continued filling of the aneurysm (R-R class 3) in 8 patients (73%). The adequate occlusion rate (R-R class 1 and 2) was 27%.

**Radiologic Outcomes at 6 Months.** On 6-month conventional angiography, complete occlusion was seen in 4 of the 11 patients (36%), with 4 (36%) having small residual necks. The remaining 3 (27%) cases showed persistent filling in the aneurysm at 6 months. The adequate occlusion rate (R-R class 1 and 2) was 72% (8 of 11 patients).

**Radiologic Outcomes at 1 Year.** One-year follow-up data were available for 9 patients. At 1 year, patient 9 improved from persistent filling (R-R class 3) to complete occlusion (R-R class 1). The remaining patients were stable with 8 of 9 patients showing adequate occlusion (R-R class 1 and 2).

**Radiologic Outcomes at 2 Years.** Two-year follow-up data were available for 9 patients. At 2 years, patient 2 improved from neck remnant (R-R class 2) to complete occlusion (R-R class 1). The remaining patients were stable with 8 of 9 patients showing adequate occlusion (R-R class 1 and 2).

**Radiologic Outcomes at 3 Years.** Three-year follow-up data were available for 8 patients. At 3-year imaging, patient 5 presented with acute SAH and angiography showed worsening of occlusion class from complete occlusion (R-R class 1) to neck filling (R-R class 2). The rupture point was from an area of regrowth along the parent vessel and under the Contour device. This patient was re-treated by using stent-assisted coiling, representing a retreatment rate of 9%. The remaining 7 patients were stable with 6 showing adequate occlusion (R-R class 1 and 2).

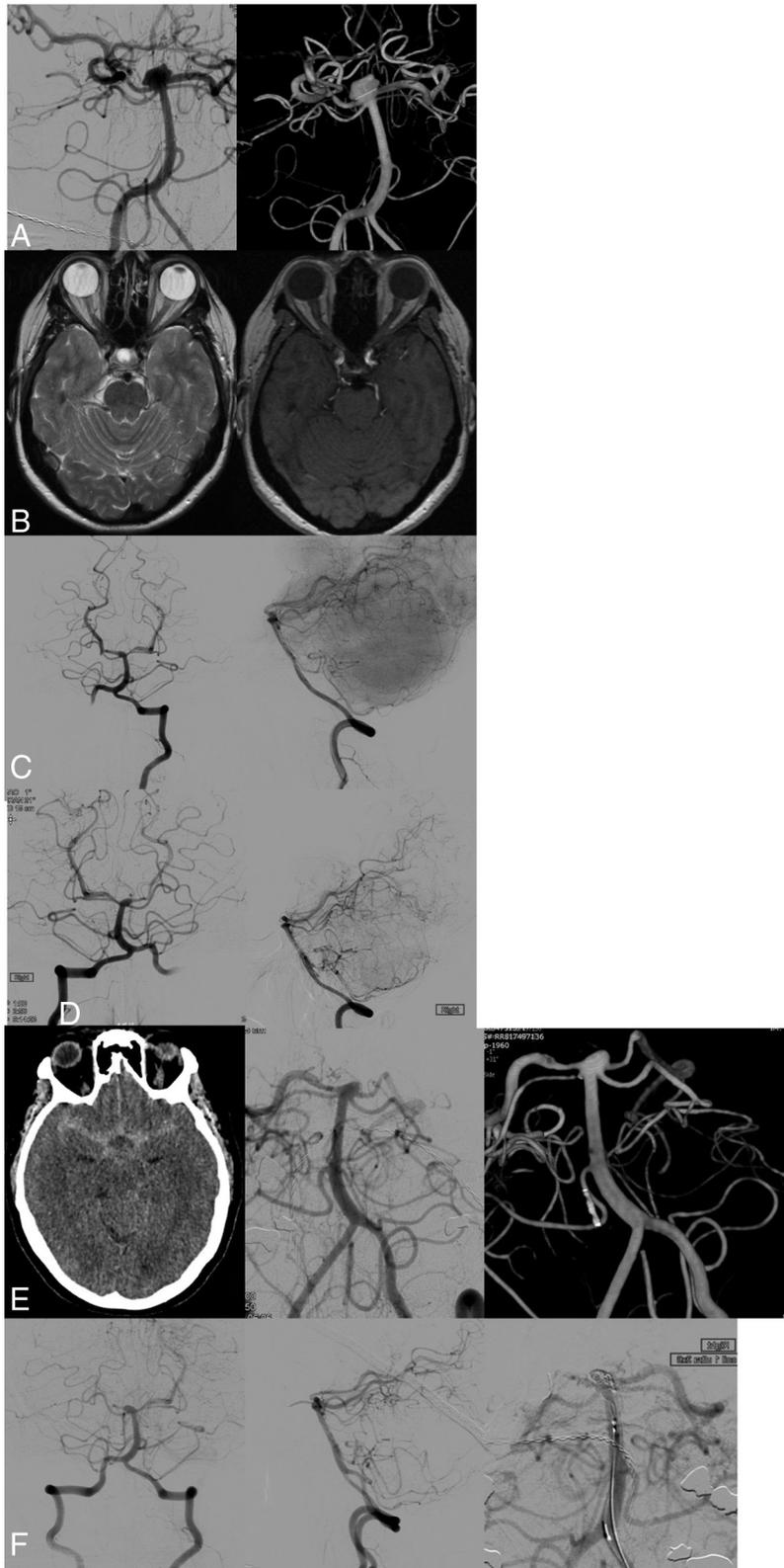
**Radiologic Outcomes at 4 Years.** Four-year follow-up data were available for 8 patients. Patient 5 had been re-treated and showed complete occlusion. Patient 8 had persistent aneurysmal filling (R-R class 3). This patient has refused further treatment. The adequate occlusion rate (R-R class 1 and 2) was 7 of 8 patients (87.5%).

**Radiologic Outcomes at 5 Years.** Five-year follow-up data were available for 7 patients. At 5-year imaging, 4 of the remaining 8 patients had stable total occlusion of the aneurysm and 2 had a stable neck remnant. Patient 3 had worsening of occlusion class, from R-R class 1 to class 2, due to growth at the neck of the aneurysm.

A summary of radiologic results is provided in Table 2. There was no evidence of device migration or compression over time.

**Table 2: Summary of radiologic results of CNS series**

Follow-Up	No. of Patients	Complete Occlusion R-R Class 1 (Number of Patients)	Small Neck Remnant R-R Class 2 (Number of Patients)	Adequate Occlusion R-R Class 1 & 2 (Number of Patients)	Residual Aneurysm R-R Class 3 (Number of Patients)
Six months	11	36% (4)	36% (4)	73% (8)	27% (3)
Year 1	9	56% (5)	33% (3)	89% (8)	11% (1)
Year 2	9	67% (6)	22% (2)	89% (8)	11% (1)
Year 3	8	50% (4)	37.5% (3)	87.5% (7)	12.5% (1)
Year 4	8	62.5% (5)	25% (2)	87.5% (7)	12.5% (1)
Year 5	7	57% (4)	43% (3)	100% (7)	0%



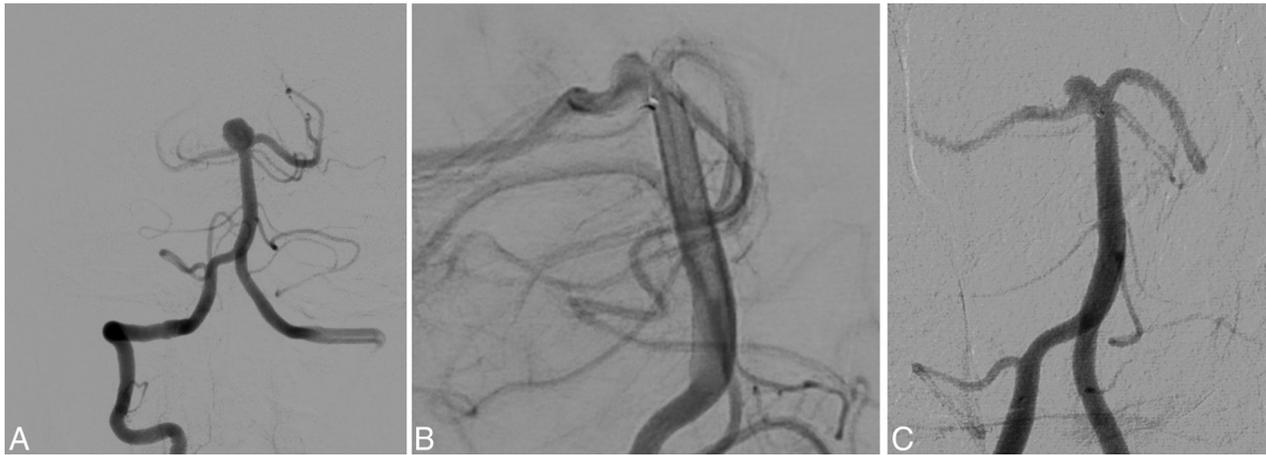
**FIG 1.** A, Pretreatment conventional angiography for patient 5, demonstrating large basilar tip aneurysm. B, MRI scan at 6-month follow-up demonstrating complete occlusion. C, Angiography at 1-year follow-up demonstrating complete occlusion. D, Angiography at 2-year follow-up demonstrating complete occlusion. E, Presentation with SAH at 3 years. F, Stent-assisted coiling treatment for acute SAH.

## DISCUSSION

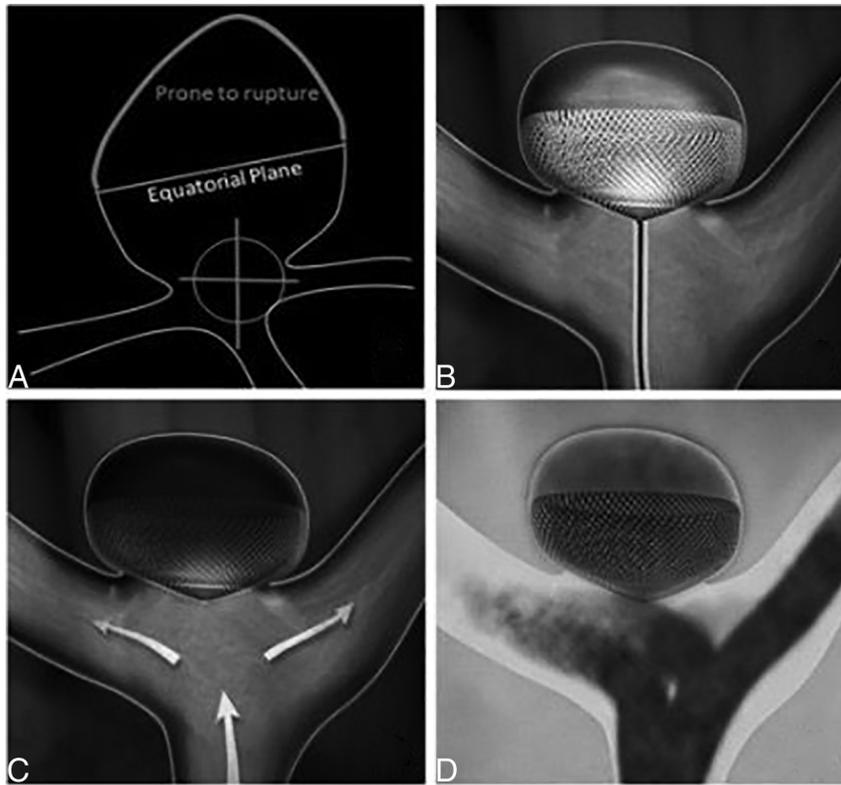
Two years ago, we published data on the first 11 patients treated with the Contour device in our center.<sup>7</sup> We continued to observe these patients to evaluate the longer-term efficacy of the device. By 4 years, we had lost 3 patients to follow-up, leaving us with 8 patients. We have at least 4 years follow-up on all remaining patients in the study (Supplemental Data and Table 2), demonstrating an adequate occlusion rate of 87.5% and stability of the device (Fig 1).

In 2021, a multicenter prospective study (Contour Neurovascular System - European Pre-Market Unruptured Aneurysm [CERUS]) evaluated the use of the Contour device for bifurcation aneurysms, and their results are very similar to ours. They demonstrated a progressive occlusion rate from 44% at 6 months to 69% at 12 months and 84% in their most recent follow-up.<sup>11</sup> They also demonstrated a thromboembolic rate of 11%, which is comparable to our thromboembolic rate of 14%. The following year, the largest series to date was published, which included 60 aneurysms treated with the Contour device.<sup>12</sup> They successfully implanted the device in 54 of 60 (90%) aneurysms and report adequate occlusion rates of 89% at 1 year<sup>12</sup> compared with our implantation success of 78% and 1-year adequate occlusion rate of 89%. For a relatively new device, it is reassuring that several studies have reported similar occlusion rates and a similar safety profile. Our series, though small, has the longest follow-up on these patients, with some interesting findings.

Although we have demonstrated satisfactory rates of occlusion, our series also includes 1 patient whose R-R class 3 filling at 6 months did not improve over 4 years, and 2 patients with worsening rates of occlusion, 1 of whom presented with acute SAH (patient 5). We had treated this patient with the Contour device for a large basilar tip aneurysm (Fig 2). The initial result was very good, with total occlusion demonstrated on MRI and conventional angiography for 2 years postprocedure. However, 3 years after treatment, the patient presented



**FIG 2.** A, Pretreatment conventional angiography for patient 3, demonstrating basilar tip aneurysm. B, Conventional angiography at 1-year follow-up, demonstrating complete occlusion of the aneurysm. C, Conventional angiography at 5-year follow-up, demonstrating recurrence at the base of the aneurysm.



**FIG 3.** A, Target zone of Contour device deployment at the intracranial aneurysm neck and within the equatorial plane. B, Schematic representation of the CNS device with catheter attached. C, Schematic representation of the CNS device deployed in the aneurysm showing expected flow diversion. D, Schematic representation of angiographic view of the CNS device.<sup>1</sup>

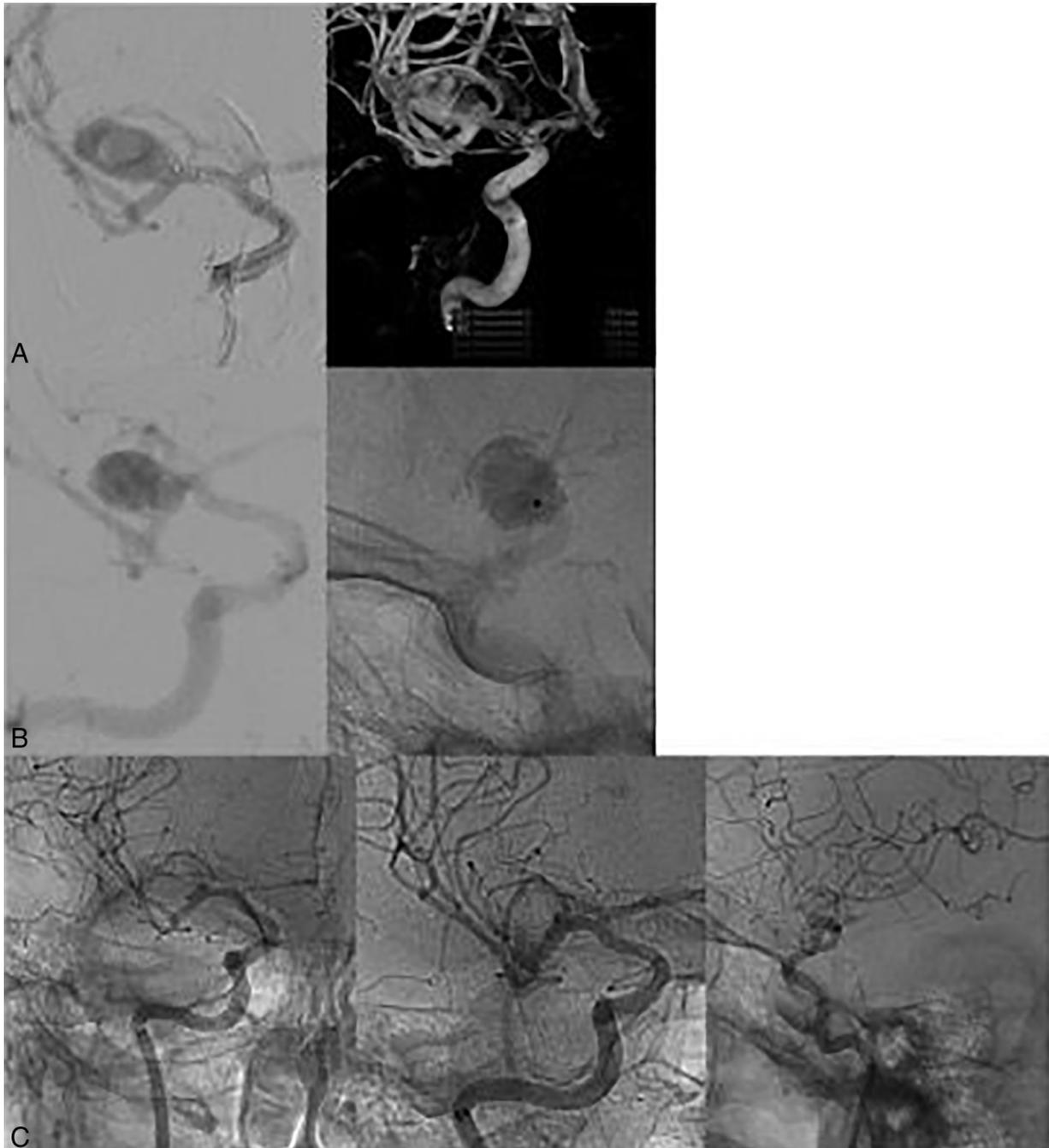
with a World Federation of Neurological Surgeons (WFNS) grade 1 SAH. Angiography demonstrated the rupture point to be from a progressive regrowth along the vessel, and under the Contour device. On re-review of angiography at 1- and 2-year follow-up, we can appreciate the dysplastic appearances of the parent vessel, something we had not appreciated at the time and a reminder of the disease process in these patients, who often fail to manage risk factors such as smoking. This a well-recognized phenomenon in aneurysmal

treatment and a similar regrowth was seen in patient 3 at 5 years (Fig 3). These 2 cases highlight to us the requirement to follow patients up, regardless of occlusion status, as regrowth at the base of the aneurysm is possible and may not be detectable for years after treatment. Furthermore, follow-up imaging should be sensitive for these regrowing aneurysms catheter and CT angiography rather than MRI that is susceptible to marker artifact at the neck.

There are several other devices currently on the market, for example the LUNA and Artisse (Medtronic), but probably the most widely used and researched is the WEB device. The WEB device is an intrasaccular embolization device that has been used increasingly in treating wide-neck aneurysms.<sup>4</sup> Several studies have demonstrated an adequate occlusion rate between 73% and 89%<sup>4,5</sup> with need for adjuvant device in 8% of cases and a thromboembolic rate of 14%.<sup>13-15</sup> This has been replicated in numerous studies.<sup>4,15-17</sup> More recently, a prospective study comparing the WEB to the Contour device demonstrated adequate occlusion rates to be similar between the 2 devices, but the rate of complete occlusion was significantly

higher for the Contour, with the WEB showing a significantly higher retreatment rate.<sup>18</sup> The results that have been published for the Contour device, by us and other groups<sup>11,12,18</sup> demonstrate that it is a good option to consider when deciding on an intrasaccular device.

The CERUS Contour study reported a need for re-treatment in 9% of patients at 1 year,<sup>11</sup> while the largest published Contour study reported a 0% re-treatment rate at 1 year.<sup>12</sup> We have demonstrated a 9% re-treatment rate at 3 years and worsening occlusion



**FIG 4.** Patient 9. *A*, Endovascular treatment of MCA aneurysm by using an 11-mm Contour device. There is complete filling of the aneurysm at the end of the procedure. *B*, Six-month follow-up angiography demonstrates continued filling of the aneurysm. *C*, One-year angiography demonstrates complete occlusion of the aneurysm, without further treatment.

rate of 18% at 5 years, further emphasizing the need to follow patients for a minimum of 5 years. Similarly, the WEB device has been associated with a re-treatment rate of 16.7%, however, these studies have also included ruptured aneurysms, making a direct comparison impossible.<sup>5,19,20</sup> Furthermore, surgical clipping continues to remain a good option, with lower rebleed and re-treatment rates than published endovascular methods.<sup>21,22</sup>

One advantage of the Contour device over the available devices is that it is comparatively easy to size and deploy. With the CNS device, only the widest diameter of the aneurysm and neck width are needed for sizing (Fig 4). Evidence in other endovascular

devices, such as the WEB device, has demonstrated better occlusion outcomes and reduced rates of compaction when oversizing the device. For that reason, we choose to oversize the Contour device when an aneurysm is suitable for 2 different sizes. In patient 8, we placed a size 11 Contour device, when a size 14 would have been more appropriate (they were not available at the time). This has resulted in persistent flow and filling of the aneurysm, even 4 years down the line.

Additionally, the Contour device allows the treatment of wide-neck aneurysms without the need for adjuvant devices due to its ability to disrupt flow in the aneurysm as well as divert flow at the neck of the aneurysm, resulting in reduced operative time.

Because it has no major parent vessel component, this removes the need for life-long antiplatelet therapy, which is a distinct advantage over devices such as the PulseRider (Cerenovus) and pCONus (phenox) and, indeed, any stent-assisted coiling technology.

One of the characteristics of the Contour device, is that it has been shown to show progressive occlusion.<sup>11</sup> We have also seen this in our cohort of patients, with aneurysms demonstrating progressive occlusion over the years. For example, in patient 9 there was continued filling of the aneurysm at 6 months. On retrospective review, we should have used a 14-mm Contour device, rather than the 11-mm device that was placed. At 1-year angiography, it was seen that the aneurysm had occluded without further treatment and the aneurysm has remained totally occluded at 5-year follow-up (Fig 4). The Contour device's progressive occlusion results from its flow-diverting properties that are like that of flow diverters.<sup>23</sup> The advantage that the Contour device has over a flow diverter is that the Contour device only needs antiplatelet treatment in the periprocedural period and up to 6–8 weeks afterward, unlike flow diverters that are often accompanied by antiplatelet therapy for years. Additionally, although flow diverters have provided an excellent option for the treatment of unruptured sidewall aneurysms, they can be ineffective for a significant percentage of aneurysms (35%–50%) that arise from bifurcations<sup>24</sup> with risk of branch occlusion, failure of aneurysm occlusion, and a situation that makes further treatment difficult.

Our initial protocol was to put all patients on a 6-week course of aspirin postoperatively, although this may need to be altered over time as more data become available. We are also considering if routine premedication with antiplatelet agents may be helpful in the run-up to the procedure and, with the failed cases, in retrospect this may have been helpful because it could have allowed the utilization of other devices in the case of failure of Contour deployment.

In our series we also found no evidence of device compression over time, which can occur with other devices.<sup>25</sup> The device position and shape remained unchanged in all cases. We believe that oversizing the device to the aneurysm is crucial to this as it subsequently prevents displacement and migration of the device into the aneurysm after deployment. In our limited experience we believe the device should not be manipulated once deployed in the aneurysm to adjust position, and if it is felt the placement is suboptimal, then it should be resheathed and redeployed in a more satisfactory position.

### Limitations

There are several limitations to this study. It is important to highlight that these were the first 11 patients we ever treated with the Contour device. A combination of familiarity and new sizing options will inevitably allow more patients to be treated with better outcomes, as seen with prior devices like the WEB.<sup>17</sup> Although we have up to 5 years of follow-up, our sample size is small and suffers from loss of patients to follow-up. It is reassuring that we demonstrate similar results to larger, multicenter studies, however, we cannot draw conclusions from such a small sample size. Second, after 2 years, most patients were followed up with MRA, as per the standard practice in our center. We appreciate that the Contour device produces pronounced

susceptibility artifacts that not only obscure the aneurysm, but also the parent vessel. This means that a remnant or recurrence may be undetectable and for that reason, most patients at 5-year follow-up had either conventional angiography or a CTA.

Of note, the efficacy of the Contour device in this study was evaluated by using the R-R classification, as the most widely used and validated scoring system. This allows comparability among endovascular techniques; however, the R-R scale was initially introduced to assess coiling outcomes. More recently the Bicêtre occlusion scale has been proposed, which provides a more comprehensive assessment of occlusion after intrasaccular device implantation and its use should be encouraged for future studies.

Finally, we have only used the Contour device in unruptured aneurysms and further studies will be essential in establishing the efficacy and safety of using the Contour device in the acute setting.

### CONCLUSIONS

In this study, we have demonstrated an adequate occlusion rate of 89% at 4 years, with evidence of occlusion being maintained at 5 years. We have demonstrated a retreatment rate of 9% with the Contour device and a 14% thromboembolic rate, which is comparable to other endovascular devices. We have also seen a worsening occlusion rate of 18% after 3 or more years of complete occlusion, emphasizing the importance of long-term radiologic follow-up in patients treated with the Contour device. Importantly, follow-up imaging needs to be sensitive for these regrowing aneurysms and MRI may be suboptimal due to artifact. Since this cohort of patients, we have continued to use the Contour device and hope to publish further data on a larger population both in our center, and in collaboration with other centers.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

### REFERENCES

1. De Leacy RA, Fargen KM, Mascitelli JR, et al. **Wide-neck bifurcation aneurysms of the middle cerebral artery and basilar apex treated by endovascular techniques: a multicentre, core lab adjudicated study evaluating safety and durability of occlusion (BRANCH).** *J Neurointerv Surg* 2019;11:31–36 CrossRef Medline
2. Fiorella D, Arthur AS, Chiacchierini R, et al. **How safe and effective are existing treatments for wide-necked bifurcation aneurysms? Literature-based objective performance criteria for safety and effectiveness.** *J Neurointerv Surg* 2017;9:1197–201 CrossRef Medline
3. Zhao B, Yin R, Lanzino G, et al. **Endovascular coiling of wide-neck and wide-neck bifurcation aneurysms: a systematic review and meta-analysis.** *AJNR Am J Neuroradiol* 2016;37:1700–05 CrossRef Medline
4. Lawson A, Goddard T, Ross S, et al. **Endovascular treatment of cerebral aneurysms using the Woven EndoBridge technique in a single center: preliminary results.** *J Neurosurg* 2017;126:17–28 CrossRef Medline
5. Clajus C, Strasilla C, Fiebig T, et al. **Initial and mid-term results from 108 consecutive patients with cerebral aneurysms treated with the WEB device.** *J Neurointerv Surg* 2017;9:411–17 CrossRef Medline
6. Piotin M, Biondi A, Sourour N, et al. **The LUNA aneurysm embolization system for intracranial aneurysm treatment: short-term, mid-term and long-term clinical and angiographic results.** *J Neurointerv Surg* 2018;10:e34 CrossRef Medline
7. Akhnbay-Fudge CY, Deniz K, Tyagi AK, et al. **Endovascular treatment of wide-necked intracranial aneurysms using the novel Contour Neurovascular System: a single-center safety and feasibility study.** *J Neurointerv Surg* 2020;12:987–92 CrossRef Medline

8. Mascitelli JR, Lawton MT, Hendricks BK, et al. **Analysis of wide-neck aneurysms in the Barrow Ruptured Aneurysm Trial.** *Neurosurgery* 2019;85:622–31 CrossRef Medline
9. Roy D, Milot G, Raymond J. **Endovascular treatment of unruptured aneurysms.** *Stroke* 2001;32:1998–204 CrossRef Medline
10. Mascitelli JR, Oermann EK, De Leacy RA, et al. **Angiographic outcome of intracranial aneurysms with neck remnant following coil embolization.** *J Neurointerv Surg* 2015;7:484–89 CrossRef Medline
11. Liebig T, Killer-Oberpfalzer M, Gal G, et al. **The safety and effectiveness of the Contour neurovascular system (Contour) for the treatment of bifurcation aneurysms: the CERUS Study.** *Neurosurgery* 2022;90:270–77 CrossRef Medline
12. Biondi A, Primikiris P, Vitale G, et al. **Endosaccular flow disruption with the Contour Neurovascular System: angiographic and clinical results in a single-center study of 60 unruptured intracranial aneurysms.** *J Neurointerv Surg* 2023;15:838–43 CrossRef Medline
13. Pierot L, Spelle L, Molyneux A, et al; WEBCAST and French Observatory Investigators. **Clinical and anatomical follow-up in patients with aneurysms treated with the WEB device: 1-year follow-up report in the cumulated population of 2 prospective, multicenter series (WEBCAST and French Observatory).** *Neurosurgery* 2016;78:133–41 CrossRef Medline
14. Pierot L, Costalat V, Moret J, et al. **Safety and efficacy of aneurysm treatment with WEB: results of the WEBCAST study.** *J Neurosurg* 2016;124:1250–56 CrossRef Medline
15. Bhogal P, Udani S, Cognard C, et al. **Endosaccular flow disruption: where are we now?** *J Neurointerv Surg* 2019;11:1024–25 CrossRef Medline
16. Fiorella D, Molyneux A, Coon A, et al; WEB-IT Study Investigators. **Demographic, procedural and 30-day safety results from the WEB Intra-saccular Therapy Study (WEB-IT).** *J Neurointerv Surg* 2017;9:1191–96 CrossRef Medline
17. Lv X, Zhang Y, Jiang W. **Systematic review of Woven EndoBridge for wide-necked bifurcation aneurysms: complications, adequate occlusion rate, morbidity, and mortality.** *World Neurosurg* 2018;110:20–25 CrossRef Medline
18. Hecker C, Broussalis E, Pfaff JAR, et al. **Comparison of the Contour Neurovascular System and Woven EndoBridge device for treatment of wide-necked cerebral aneurysms at a bifurcation or sidewall.** *J Neurosurg* 2023;139:563–72 CrossRef Medline
19. Lubicz B, Klisch J, Gauvrit JY, et al. **WEB-DL endovascular treatment of wide-neck bifurcation aneurysms: short- and midterm results in a European study.** *AJNR Am J Neuroradiol* 2014;35:432–38 CrossRef Medline
20. Pierot L, Klisch J, Liebig T, et al. **WEB-DL endovascular treatment of wide-neck bifurcation aneurysms: long-term results in a European series.** *AJNR Am J Neuroradiol* 2015;36:2314–19 CrossRef Medline
21. Kim YD, Bang JS, Lee SU, et al. **Long-term outcomes of treatment for unruptured intracranial aneurysms in South Korea: clipping versus coiling.** *J Neurointerv Surg* 2018;10:1218–22 CrossRef Medline
22. Molyneux A, Kerr R, Stratton I, et al; International Subarachnoid Aneurysm Trial (ISAT) Collaborative Group. **International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised trial.** *Lancet* 2002;360:1267–74 CrossRef Medline
23. Bhogal P, Martinez Moreno R, Ganslandt O, et al. **Use of flow diverters in the treatment of unruptured saccular aneurysms of the anterior cerebral artery.** *J Neurointerv Surg* 2017;9:283–89 CrossRef Medline
24. Bhogal P, AlMatter M, Hellstern V, et al. **Difference in aneurysm characteristics between ruptured and unruptured aneurysms in patients with multiple intracranial aneurysms.** *Surg Neurol Int* 2018;9:1 CrossRef Medline
25. Cognard C, Januel AC. **Remnants and recurrences after the use of the WEB intrasaccular device in large-neck bifurcation aneurysms.** *Neurosurgery* 2015;76:522–30 CrossRef Medline

# Safety and Effectiveness of a Novel Integrated Angiography System for Continuous Guidance and Flushing in Diagnostic Cerebral Catheter Angiography: A Randomized Controlled Trial

 Boseong Kwon, Jong-Tae Yoon,  Yun Hyeok Choi,  Soo Jeong, Byung Jun Kim, Joon Ho Choi, Sun Moon Hwang,  Deok Hee Lee, and  Yunsun Song



## ABSTRACT

**BACKGROUND AND PURPOSE:** The guided angiography system, an integrated angiography system, enables continuous guidance and flushing during diagnostic procedures. A guidewire provides guidance by remaining inside the catheter during contrast injection into tortuous vessels. While its feasibility was demonstrated in a retrospective study, this randomized controlled trial aims to evaluate its safety and effectiveness.

**MATERIALS AND METHODS:** This single-center, single-blind, randomized controlled trial was conducted from September 2021 to June 2022. Patients with unruptured intracranial aneurysms were enrolled and randomly assigned to either the guided or the conventional group. Primary outcomes were procedural time and serious adverse events. Secondary outcomes included catheterization success, adverse events, fluoroscopy time, fluoroscopy dose, and image quality.

**RESULTS:** A total of 200 participants (mean age, 59 [SD, 10 years], 71% women) were randomized into the guided ( $n=100$ ) and the conventional ( $n=100$ ) groups. The guided group demonstrated shorter procedural times compared with the conventional group (18.3 [SD, 9.2] versus 21.3 [SD, 8.1] minutes,  $P < .001$ ). There were no serious adverse events in either group, and adverse events were similar. Catheterization success rates were similar between the guided and conventional groups (93.9% versus 96.8%,  $P = .226$ ). However, the guided angiography system backed up more catheterization failures from the conventional angiography system, but not vice versa (57.1% versus 11.1%,  $P = .003$ ). Among catheterization successes, the guided group achieved shorter fluoroscopy time (7.0 versus 7.7 minutes,  $P = .033$ ) and lower fluoroscopy dose (466.8 versus 566.5  $\mu\text{Gy}\cdot\text{m}^2$ ,  $P = .015$ ). Regarding imaging quality, clinical acceptability was comparable between the guided and conventional groups (95.3% versus 95.5%,  $P = .781$ ).

**CONCLUSIONS:** The guided angiography system offers a simple and effective alternative for diagnostic cerebral conventional angiography, with reduced radiation exposure and a favorable safety profile.

**ABBREVIATION:** AE = adverse event

Cerebral conventional angiography plays a crucial role in diagnosing and evaluating various neurovascular diseases. However, this procedure is invasive and could result in neurologic complications, with reported rates of up to 2.5%.<sup>1-3</sup> The conventional diagnostic catheters used are typically 4Fr or 5Fr with inner diameters of 0.041 or 0.043 inches, paired with 0.035- and 0.038-inch guidewires.<sup>4,5</sup>

These configurations require the removal of the guidewire whenever a contrast agent injection is performed, necessitating additional steps to ensure safety. Furthermore, the presence of thrombi or unintentionally introduced air bubbles in the catheter can lead to thromboembolic complications.<sup>6-9</sup> Despite advancements in various neurointerventional devices, considerable improvements have not been made to the diagnostic angiography system.

To address these issues, we developed the guided angiography system, an integrated angiography system, comprising a 5Fr catheter with an expanded inner diameter of 0.054 inches, a 0.032-inch fortified guidewire, a hemostatic valve, and a manifold. This

Received September 23, 2024; accepted after revision February 3, 2025.

From the Department of Radiology (B.K., D.H.L., Y.S.), Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; Department of Radiology (J.H.C., S.M.H.), Asan Medical Center, Seoul, South Korea; Biomedical Engineering Research Center (J.-T.Y.), Asan Institute for Life Sciences, Asan Medical Center, Seoul, South Korea; Department of Neurosurgery (Y.H.C.), Inje University Haeundae Paik Hospital, Busan, South Korea; Department of Neurology (S.J.), Hanyang University Medical Center, Seoul, South Korea; and Advanced Therapies (B.J.K.), Siemens Healthineers Co. Ltd., Seoul, South Korea.

Boseong Kwon and Jong-Tae Yoon contributed equally to the study as first authors.

This work was supported by the "Supporting Project to Evaluation New Domestic Medical Devices in Hospitals" funded by the Ministry of Health and Welfare and the Korea Health Industry Development Institute, which had no role in the design, analysis, interpretation, or publication of this study.

Please address correspondence to Yunsun Song, MD, PhD, Department of Radiology, Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, 88, Olympic-ro 43-gil, Songpa-gu, Seoul, 05505, South Korea; e-mail: songsongggg@gmail.com



This article contains an imaging protocol available on our website.

<http://dx.doi.org/10.3174/ajnr.A8700>

## SUMMARY

**PREVIOUS LITERATURE:** In a conventional cerebral angiography system, a continuous flushing system cannot be applied in traditional angiography practice due to the limited space between the catheter and the guidewire. Repeat maneuvers, such as intermittent flushing and ensuring air-free connections in the lines, inherently carry a risk of embolism, partly due to the potential for human error during these processes. Furthermore, the lack of guidewire guidance before contrast injection occasionally results in catheter displacement, necessitating re-catheterization or re-acquisition of images.

**KEY FINDINGS:** In this randomized study, the guided angiography system demonstrated advantages over the conventional angiography system, including shorter procedural times, improved convenience, better catheterization in difficult cases, and reduced fluoroscopy time and radiation dose, while maintaining safety and clinically acceptable image quality despite slight reductions.

**KNOWLEDGE ADVANCEMENT:** The integration of device advancements securing flow space between the catheter and guidewire, manifolds with flushing and contrast medium lines, and a hemostatic valve enables continuous guidance and flushing during diagnostic cerebral conventional angiography. This integrated angiography system has the potential to enhance the efficiency and safety of diagnostic procedures.

novel angiography system renders flow space between the catheter and guidewire for both continuous flushing and intermittent contrast injection while eliminating the need to remove the guidewire. Additionally, the system benefits from continuous guidance due to the full-length presence of the guidewire throughout the procedure. Its feasibility and usefulness in diagnostic cerebral conventional angiography have been demonstrated in previous studies.<sup>10,11</sup> Potential advantages of this novel angiography system include a streamlined procedure, enhanced catheterization by continuous guidance, and a minimized risk of thrombus formation and air entry by a closed and continuous flushing system.

We hypothesize that the guided angiography system provides a more effective alternative to the conventional angiography system, while maintaining the safety profile. To test this hypothesis, we conducted a randomized controlled trial. This article follows the CONSORT Reporting Guidelines (<https://pmc.ncbi.nlm.nih.gov/articles/PMC6398298/>).

## MATERIALS AND METHODS

### Study Design

This single-blind, randomized, controlled clinical trial was conducted at Asan Medical Center, Seoul, South Korea. The research

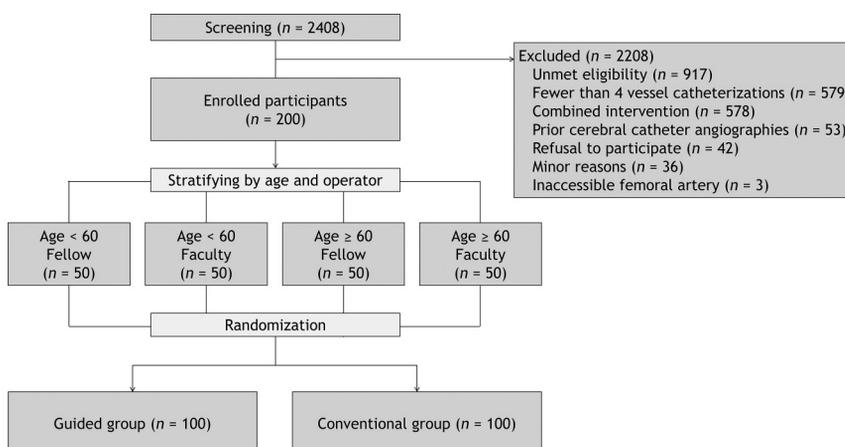
protocol and informed consent forms were approved by the institutional review board (2021–2022) and registered with the Clinical Research Information Service (Cris.nih.go.kr; KCT0006644). Written informed consent was obtained from all participants before enrollment.

### Participants and Randomization

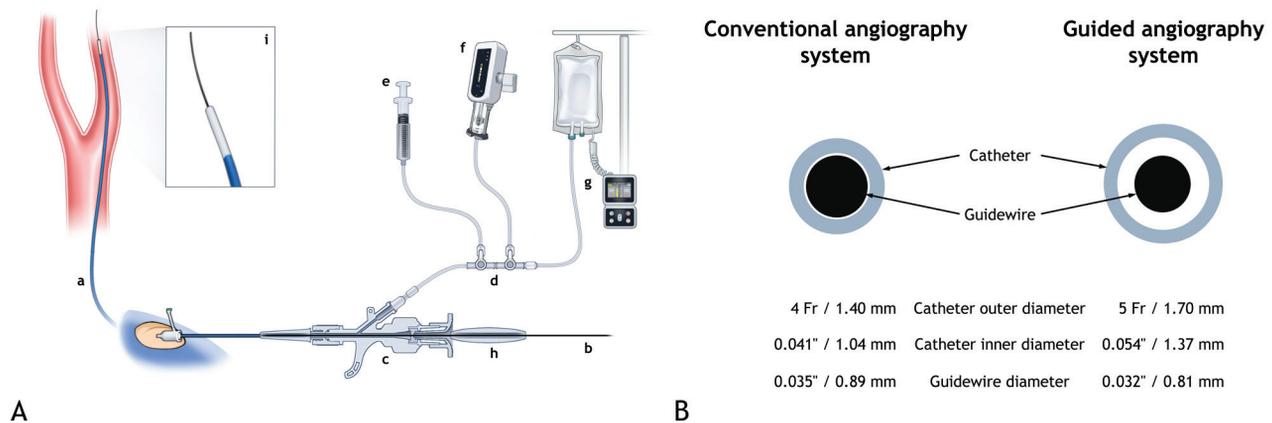
The study included eligible participants with unruptured intracranial aneurysms who were 18 years of age or older and scheduled to undergo diagnostic cerebral conventional angiography (Fig 1). Exclusion criteria included an inaccessible femoral artery, previous angiography, combined procedures, examination of fewer than 4 vessels, or unwillingness to participate in the trial. Participants were randomized in a 1:1 ratio stratified by age (younger than 60 years of age or 60 years of age and older) and operator (fellow or faculty) into either the guided angiography system (guided group) or the conventional angiography system group (conventional group). A statistician generated the randomization sequence using a permuted block randomization scheme. The variable blocks with the sizes of 4 and 6 were randomly ordered. When a participant was enrolled by investigators, an independent research assistant performed study group allocation referring to the randomization sequence. All participants and investigators, except for the operators, were blinded to the group assignments.

### Angiography Systems

The guided angiography system secures a certain amount of space between the catheter and guidewire by increasing the catheter lumen and decreasing the need for guidewire removal for contrast agent injection. A hemostatic valve is included to prevent blood regurgitation, and a manifold provides a hub for multiple lines. A torque device was used to secure the guidewire during contrast medium injection. Detailed information can be found in Fig 2A, -B



**FIG 1.** Flowchart of patient enrollment and randomization.



**FIG 2.** A, Schematic representation of the components of the integrated angiography system with continuous guidance and flushing. This system comprises the following: a) a 5Fr 100-cm angiography catheter (GRAFIA; Sungjin-Hitech) with a 0.054-inch inner diameter; b) a 0.032-inch 150-cm guidewire (ANGUIS; Sungjin-Hitech); c) a hemostatic valve; and d) a manifold that receives lines from a syringe (e), a power injector (f), and a pressurized saline bag (g) (P-pump; Impact Korea). Continuous heparinized saline flush prevents blood backflow during catheter navigation. When the catheter reaches the target position, the power injector line is opened, and the guidewire is fixed using the torque device (h). In the contrast agent injection step, the guidewire remaining in the catheter stabilizes the catheter tip and guides it along the axis of the catheterized vessel (i). The syringe (e) is used for initial catheter preparation and manual roadmap acquisition. B, Comparison of the cross-sectional dimensions of the conventional angiography system and the integrated angiography system with continuous guidance and flushing (guided angiography system).

and previous reports.<sup>10,11</sup> The conventional angiography system consisted of a 4Fr diagnostic catheter (JSM angio catheter; Jung Sung Medical) and a 0.035-inch guidewire (Glidewire; Terumo Medical). For both angiography systems, a simple 45° catheter tip shape was used.

### Angiography Protocols

A biplane angiography machine (Artis zee biplane; Siemens) was used for image and data acquisition. A 4Fr or 5Fr introducer sheath was selected on the basis of the assigned angiography system. Bilateral internal carotid and vertebral angiographies were obtained. External carotid and 3D rotational angiographies were obtained if necessary. If the assigned angiography system failed, the other angiography system was used as a backup. If both angiography systems failed, a reverse-curve 4Fr diagnostic catheter was used as a rescue to complete the examination. All angiography procedures were performed in the usual manner, except for keeping the guidewire inside the catheter and securing it in place before contrast injection when using the guided angiography system. During the examination of tortuous vessels with the guided angiography system, continuous guidance was used at the operator's discretion to guide and maintain the catheter tip position by advancing the guidewire a few centimeters beyond the catheter tip. The contrast agent (Pamiray 300; Dongkook Pharmaceutical) and the injection protocol were identical for both groups. Hemostasis was achieved using a compressive hemostatic device.

### Assessments and Outcomes

We assessed patient characteristics, vascular risk factors, and adverse events (AEs), including neurologic deficit, vascular injury (iatrogenic dissection and flow-limiting vasospasm), intraprocedural thromboembolism, contrast-induced complications, and puncture site complications (hematoma, pseudoaneurysm, and arteriovenous fistula). AEs were evaluated on the examination day and again after 2 weeks. The primary effectiveness outcome

was procedural time, defined as the duration from femoral puncture to the completion of the examination, including the system preparation time. Procedural time was used as a surrogate outcome for a streamlined procedure and enhanced catheterization. The primary safety outcome was a composite of serious AEs that extended the hospitalization period or required re-admission. Secondary outcomes included catheterization success, any AEs, fluoroscopy time, and fluoroscopy dose. Catheterization success was defined as the successful catheterization of each vessel using only the assigned system and was used as a measure of technical success. Fluoroscopy time and dose were provided in the dose report.

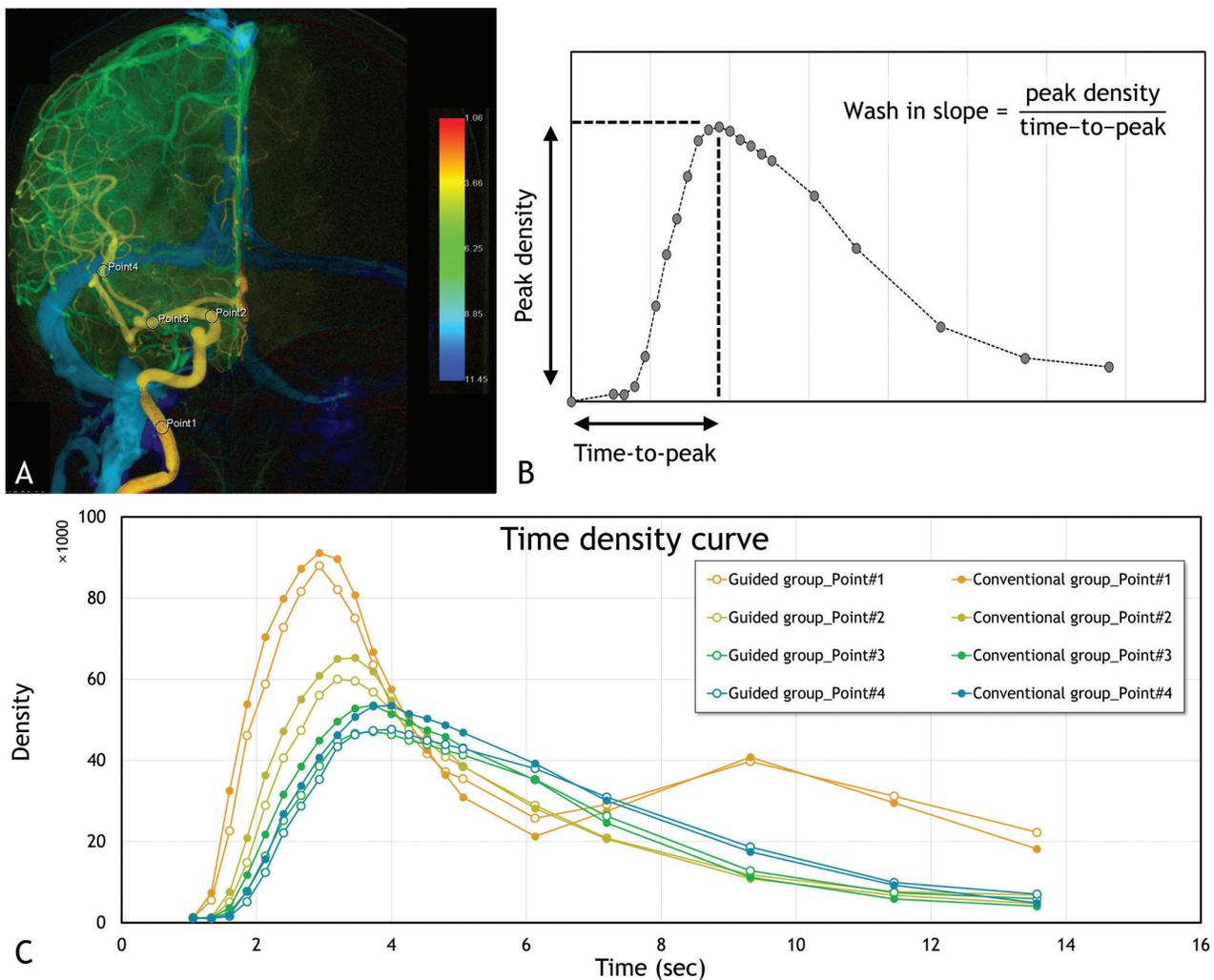
### Image-Quality Evaluation

DSA image quality was evaluated using quantitative and qualitative metrics. Quantitative analysis was performed using specialized software (syngo iFlow; Siemens) to generate time-density curves and to compare peak attenuation, time-to-peak, and wash-in flow slope at 4 regions of interest between groups (Fig 3A, -B).

Qualitative evaluation was performed by 3 neuroradiologists who assessed the images using a 5-point Likert scale, for which a score of  $\geq 3$  was deemed clinically acceptable. The rating scale was as follows: "excellent" (5-point), indicating sufficient attenuation in all vessels; "good" (4-point), indicating insufficient attenuation of some vessels but not markedly hindering the evaluation of vascular lesions; "fair" (3-point), indicating low overall attenuation that limits the assessment of small distal vessels; "poor" (2-point), indicating markedly low overall attenuation that restricted the evaluation of overall vessels; and "fail" (1-point), indicating markedly poor overall attenuation that necessitates re-examination.

### Statistical Analysis

On the basis of the pilot study,<sup>10</sup> we assumed the mean (SD) to be 21.5 and 7.0 in the guided group and 25.1 and 8.3 in the conventional group, respectively. To detect a significant difference between the 2 groups, we calculated a sample size of 100



**FIG 3.** Quantitative analysis using iFlow. **A**, Time-density curves are analyzed in 4 regions of interest: the proximal ICA (point 1), distal ICA (point 2), M1 segment (point 3), and M2 segment (point 4). **B**, Peak attenuation is defined as the maximum intensity throughout the angiography acquisition; time-to-peak is defined as the duration to reach maximum intensity from the start of contrast injection. The wash-in flow slope is defined as the peak attenuation divided by time-to-peak. **C**, The average time-attenuation curves of both groups in each region of interest.

participants per group, considering a 2-sided significance level of .05, a power of 0.9, and a drop-out rate of 5%. Outcomes were primarily analyzed on an intention-to-treat basis. To precisely evaluate the effect of the sole angiography system, we additionally analyzed the outcomes of participants who did not undergo any backup or rescue attempts. Continuous variables are presented as mean (SD), and categorical variables are presented as frequencies with percentages. Continuous variables were analyzed using the Student *t* test or Wilcoxon rank-sum test, while categorical variables were analyzed using the  $\chi^2$  test or Fisher exact test. Stratified analysis was conducted using analysis of covariance after log transformation. Catheterization success and qualitative image-quality analysis were performed using a log-linear model with a generalized estimating equation on a per-vessel basis. Interrater reliability was assessed using the Gwet AC2 agreement coefficient. A 2-sided significance level of  $<.05$  was considered statistically significant. All statistical analyses were conducted using SAS (Version 9.4; SAS Institute,) or R (Version 3.6.1; <http://www.r-project.org/>).

## RESULTS

A total of 200 participants 59 (SD, 10) years of age (women,  $n = 142$  [71%]) were enrolled between September 2021 and June 2022 and were randomly assigned to either the guided group ( $n = 100$ ) or the conventional group ( $n = 100$ ). The baseline characteristics of both groups showed no significant differences (Table 1). In each of the 2 groups, 100 examinations were completed, and 441 vessels were catheterized. The catheterization success was 93.9% (414/441 vessels) in the guided group and 96.8% (427/441 vessels) in the conventional group, which were not significantly different ( $P = .226$ ). During the backup attempts, the guided angiography system achieved significantly more successes than the conventional angiography system (57.1% versus 11.1%,  $P = .003$ ) (Fig 4).

The procedural time was significantly shorter in the guided group compared with the conventional group, with a reduction of 14.1% (18.3 [SD, 9.2] versus 21.3 [SD, 8.1] minutes;  $P < .001$ ) (Table 2 and Fig 5A). In a stratified analysis with a cutoff at 60 years of age, the guided group had significantly shorter

procedural times compared with the conventional group in both age strata (younger than 60 years: 14.8 versus 18.1 minutes;  $P = .001$ ; 60 years of age and older: 21.8 versus 24.6 minutes;  $P = .009$ ) (Fig 5B). When stratified by the level of operator experience (ie, fellow versus faculty), the conventional group had a

significantly longer procedural time than the guided group in the fellow stratum (22.5 versus 18.0 minutes,  $P < .001$ ), while the difference was not statistically significant in the faculty stratum (20.2 versus 18.7 minutes,  $P = .098$ ) (Fig 5C). There was no interaction between the strata and the study groups ( $P = .660$  for age,  $P = .174$  for operator experience).

No serious AEs were observed in either group. The guided group had 2 cases of contrast-induced AEs, while the conventional group had 5 cases. Puncture site hematomas were reported in 26 participants in the guided group and 22 participants in the conventional group, but this difference was not statistically significant.

The fluoroscopy time and fluoroscopy dose did not differ significantly between the 2 groups (8.7 [SD, 6.9] versus 8.8 [SD, 5.8] minutes,  $P = .152$ ; 630.4 [SD, 694.2] versus 609.4 [SD, 432.1]  $\mu\text{Gy}\cdot\text{m}^2$ ,  $P = .136$ ) (Table 2).

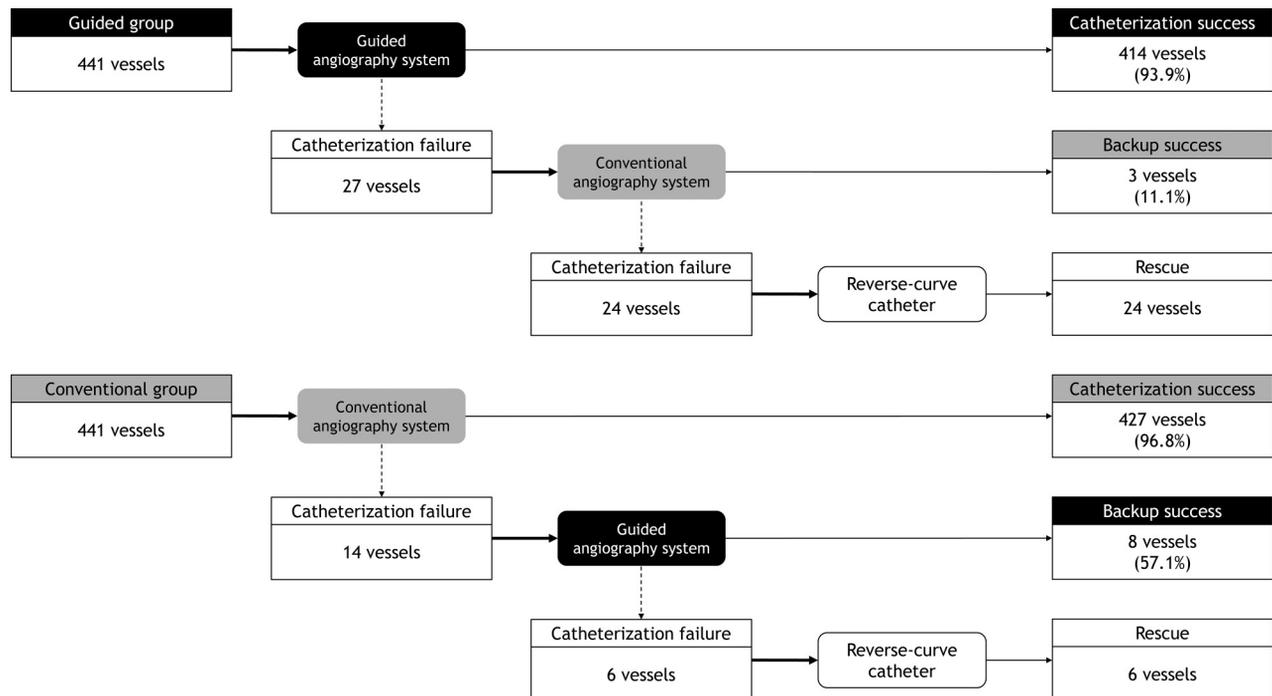
An additional analysis excluding participants who underwent any backup or rescue attempts was conducted to evaluate the effects of the sole angiography system (Table 2). The procedural time was even shorter in the guided group compared with the conventional group, showing a reduction of 25.1% (15.9 [SD, 4.2] versus 19.8 [SD, 4.9] minutes,  $P < .001$ ). Additionally,

**Table 1: Baseline characteristics<sup>a</sup>**

Characteristics	Guided Group (n = 100)	Conventional Group (n = 100)	P Value
Age (yr)	59.4 (SD, 9.7)	58.8 (SD, 10.3)	.687
Female sex	73 (73)	69 (69)	.533
Risk factors			
Hypertension	42 (42)	44 (44)	.775
Hyperlipidemia	14 (14)	18 (18)	.440
Diabetes mellitus	13 (13)	13 (13)	1.000
Smoking	12 (12)	21 (21)	.086
Alcohol consumption	27 (27)	31 (31)	.533
Aortic arch type			.437
Type I	65 (65)	70 (70)	
Type II	24 (24)	24 (24)	
Type III	11 (11)	6 (6)	
Bovine arch variant	6 (6)	5 (5)	.756
No. of vessels examined	4.5 (SD, 0.7)	4.4 (SD, 0.6)	.548
No. of 3D-RA performed	2.3 (SD, 1.0)	2.5 (SD, 1.2)	.334
Aneurysm distribution			.091
ICA	68 (40.7)	56 (33.9)	
ACA	44 (26.3)	42 (25.5)	
MCA	49 (29.3)	50 (30.3)	
Posterior circulation	6 (3.6)	17 (10.3)	
Total	167 (100)	165 (100)	

**Note:**—ACA indicates anterior cerebral artery; 3D-RA, 3D rotational angiography.

<sup>a</sup> Values are presented as mean (SD) or number (%).



**FIG 4.** Flowchart of the examination process for the 2 groups. Catheterization was assessed on a per-vessel basis. For both angiography systems, a simple 45° catheter tip shape was used. Catheterization success is defined as successful catheterization of a target vessel using the assigned angiography system. Backup success refers to successful catheterization using the other angiography system once the assigned angiography system failed. Rescue involves catheterization using a reverse-curve catheter when both angiography systems fail. The guided angiography system refers to the integrated angiography system with continuous guidance and flushing.

the fluoroscopy time was significantly reduced in the guided group (7.0 [SD, 3.5] versus 7.7 [SD, 3.3] minutes,  $P = .033$ ), and the fluoroscopy dose was significantly lower in the guided group (466.8 [SD, 298.3] versus 566.5 [SD, 359.0]  $\mu\text{Gy}\cdot\text{m}^2$ ,  $P = .015$ ).

Regarding quantitative image analysis, there were significant differences between the groups in terms of peak intensity at points 3 (52.7 versus 58.5,  $P = .001$ ) and 4 (53.4 versus 58.9,  $P = .001$ ), as well as the wash-in flow slope at point 3 (14.3 versus 16.1,  $P < .001$ ) (Table 3 and Fig 3C). In contrast, there were no significant differences between the groups in terms of the time-to-peak or peak intensity at points 1 and 2.

A significant difference was found between the 2 groups in terms of qualitative image quality, with the conventional group having a higher frequency of higher grades (4 and 5) compared with the guided group (77.6% versus 83.7%,  $P < .001$ ). However, in terms of an acceptable image-quality score of  $\geq 3$ , there was no significant difference between the guided group and the conventional group (95.3% versus 95.5%,  $P = .781$ ). The interrater agreement was good ( $\kappa = 0.826$ ) (Table 3).

## DISCUSSION

This study demonstrated that the use of the guided angiography system resulted in a shorter procedural time with comparable catheterization success compared with the conventional angiography system. No serious AEs were reported in either group. The image quality obtained in the guided group was acceptable and comparable with that of the conventional group.

The primary effectiveness outcome of this study—procedural time—consists of the cumulative durations of 2 repeat steps: the catheterization step, which involves the manipulation of the catheter and guidewire under fluoroscopic guidance, and the contrast-injection step, which includes processes such as guidewire removal, blood aspiration, line connection for contrast injection (or hand injection), and guidewire re-insertion. Considering that there was no significant difference in catheterization success, the simplification of the contrast-injection step offered by the guided angiography system might play an important role in the difference in the procedural time (Fig 2C, -D). When it comes to the

longer conventional procedural times in the fellow stratum, we believe that the guided angiography system contributes to user convenience and provides easier learning for novice operators. To be more specific, the laborious contrast-injection step mandatory in the conventional angiography system likely contributed to procedural burdens, especially for less experienced operators.

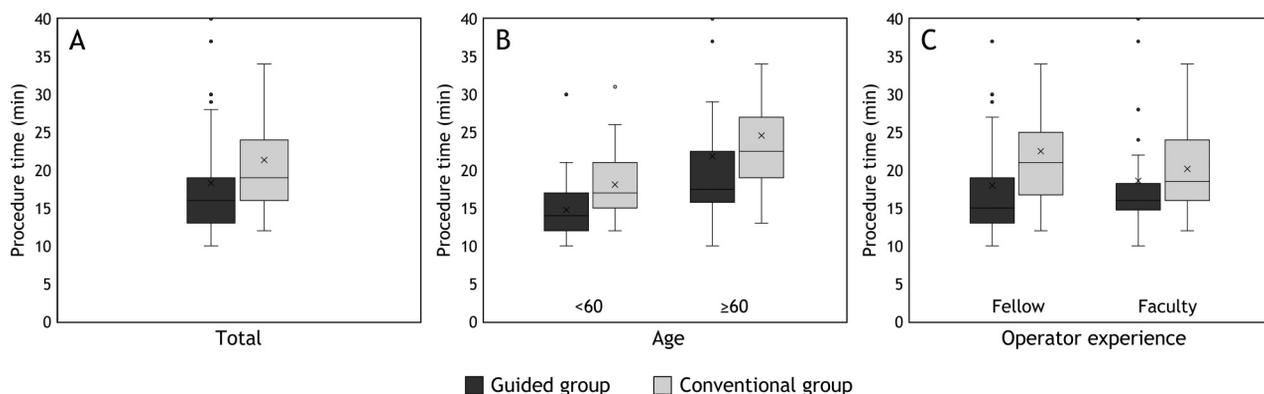
Catheterization success depends on the navigability, supportiveness, and trackability of the angiography system, which are necessary to overcome arterial tortuosity and achieve the desired catheter tip position. More successful and higher-quality catheterization results in less fluoroscopy operation, potentially leading to reduced fluoroscopy time and a lower fluoroscopy dose. In this regard, the guided angiography system may have device deficiencies due to its thinner catheter wall and guidewire, which are necessary to secure the flowable space

**Table 2: Primary and secondary outcomes<sup>a</sup>**

	Guided Group	Conventional Group	P Value
Procedural time (min)			
All	18.3 (SD, 9.2)	21.3 (SD, 8.1)	<.001
Excluding backup/rescue	15.9 (SD, 4.2)	19.8 (SD, 4.9)	<.001
Stratified analysis			
Age			.660
<60 years	14.8 (SD, 3.5)	18.1 (SD, 4.3)	.001
$\geq 60$ years	21.8 (SD, 11.5)	24.6 (SD, 9.7)	.009
Operator experience			.174
Fellow	18.0 (SD, 9.0)	22.5 (SD, 9.4)	<.001
Faculty	18.7 (SD, 9.6)	20.2 (SD, 6.5)	.098
Serious AEs	0	0	
AEs			
Contrast-induced	2 (2)	5 (5)	.445
Puncture site hematoma	26 (26)	22 (22)	.619
Catheterization success	414/441 (93.9)	427/441 (96.8)	.226
Fluoroscopy time (min)			
All	8.7 (SD, 6.9)	8.8 (SD, 5.8)	.152
Excluding backup/rescue	7.0 (SD, 3.5)	7.7 (SD, 3.3)	.033
Fluoroscopy dose ( $\mu\text{Gy}\cdot\text{m}^2$ )			
All	630.4 (SD, 694.2)	609.4 (SD, 432.1)	.136
Excluding backup/rescue	466.8 (SD, 298.3)	566.5 (SD, 359.0)	.015

**Note:**—min indicates minute.

<sup>a</sup> Values are presented as mean (SD) or number (%).



**FIG 5.** Comparison of the primary effectiveness outcome between the guided and conventional groups across total participants (A), age strata (B), and operator experience strata (C).

**Table 3: Image-quality evaluation<sup>a</sup>**

	Guided Group	Conventional Group	P Value
Quantitative analysis			
Peak intensity			
Point 1	99.0 (SD, 23.8)	100.9 (SD, 23.5)	.602
Point 2	68.5 (SD, 13.9)	72.5 (SD, 15.6)	.074
Point 3	52.7 (SD, 10.0)	58.5 (SD, 12.2)	.001
Point 4	53.4 (SD, 9.7)	58.9 (SD, 13.0)	.001
TTP (sec)			
Point 1	3.09 (SD, 0.49)	3.11 (SD, 0.44)	.659
Point 2	3.40 (SD, 0.51)	3.47 (SD, 0.45)	.334
Point 3	3.78 (SD, 0.71)	3.67 (SD, 0.51)	.276
Point 4	4.04 (SD, 0.96)	4.09 (SD, 0.68)	.643
Wash-in flow slope			
Point 1	32.7 (SD, 8.9)	33.0 (SD, 8.6)	.858
Point 2	20.5 (SD, 4.9)	21.2 (SD, 4.9)	.358
Point 3	14.3 (SD, 3.2)	16.1 (SD, 3.4)	<.001
Point 4	13.8 (SD, 3.7)	15.0 (SD, 4.2)	.090
Qualitative analysis			<.001
5	537 (40.6)	727 (55.0)	
4	490 (37.0)	379 (28.7)	
3	234 (17.7)	158 (11.9)	
2	50 (3.8)	46 (3.5)	
1	12 (0.9)	13 (1.0)	
Acceptable image quality ( $\geq 3$ )	95.3%	95.5%	.781

<sup>a</sup> Values are presented as mean (SD) or as number (%).

for the contrast agent. Although both fluoroscopy time and dose were not significantly different between the 2 groups in the intention-to-treat analysis, they were significantly different in the analysis of the cohort that excluded participants who underwent any backup or rescue attempts. Considering similar catheterization success in both groups, we believe that reduced fluoroscopy time and a lower fluoroscopy dose in the guided group without backup or rescue result from the enhanced performance of the guided angiography system. In addition, the guided angiography system successfully backed up catheterization failures that occurred with the conventional angiography system, with statistical significance. Our results suggest that the mechanical properties of the catheter and guidewire used in the guided angiography system are not inferior to those of the conventional angiography system, despite having a thinner catheter wall and guidewire.

On the basis of the literature, the guided angiography system may potentially reduce thromboembolic complications by inhibiting thrombus formation and preventing air entry, achieved through continuous flushing with pressurized heparinized saline and the use of a closed manifold system.<sup>4,5,12</sup> In contrast, the conventional catheter system requires particular attention to the double flushing technique and meticulous air-free line connections to achieve the same purpose, leading to procedural burdens and increased complexity. No serious AEs were observed in either group, which hindered the comparison of the primary safety outcome. However, this outcome suggests that both systems can be safely used in cerebral angiography when used appropriately, particularly in high-volume neurointervention centers (>2000 cases/year).<sup>12</sup> A future study comparing the safety outcomes of the 2 systems in low-volume centers or those with a relatively high rate of neurologic complications using the conventional angiography system may be helpful.

Iatrogenic dissection and vasospasm are uncommon complications related to the guidewire, catheter tip, or contrast injection flow directed toward the intima.<sup>1-3,13</sup> However, none of these

complications were observed in either group. Proper catheter tip positioning is crucial to minimize such complications.<sup>5</sup> In the guided angiography system, the direction of the catheter tip can be constantly aligned with the catheterized vessel with the help of the guidewire, which is slightly advanced in the catheterized vessel, even during the contrast-injection period. In contrast, in the conventional system, the catheter tip position may shift unfavorably following guidewire withdrawal, potentially contributing to vasospasm or even dissection.

Regarding image-quality analysis, the guided group showed lower peak intensity and slope at some points in the time-density curve. Although there was no significant difference in the clinical acceptability of image quality according to the assessment of the raters, there was a significant difference in the distribution of scores, which may be explained as follows.

The cross-sectional area between the catheter and guidewire in the guided angiography system is 0.958 mm<sup>2</sup>, which is larger than the inner lumen of the conventional angiography system (0.849 mm<sup>2</sup>). However, the contact area with the contrast agent more than doubles. Consequently, even using the low-viscosity contrast agent, there is a considerable increase in resistance against the contrast agent flow during contrast injection. This results in attenuated rise of the flow rate with a slightly longer injection time for the completion of the determined contrast volume injection, leading to a decrease in image quality. These issues could potentially be resolved by both increasing the flow rate to offset of the attenuated rise and decreasing the resistance by improving the system. Further research is needed to address these matters.

This study has several limitations that should be noted. First, the number of participants was insufficient to detect differences in safety outcomes between the groups due to the low complication rate. However, the sample size was considered adequate for evaluating the primary effectiveness outcome. Another issue related to the study population is that we would include fewer patients with atherosclerosis. A study performed in a cohort with more patients with atherosclerosis may provide informative results. Second, the conventional group used a 4Fr catheter with a simple 45° curve. It could be argued that comparing the conventional 4Fr catheter with the 5Fr catheter in the guided angiography system is inappropriate because the conventional 5Fr catheter may have superior mechanical properties compared with the conventional 4Fr catheter. However, the 4Fr catheter had already been used as part of our best practices at our center; therefore, we believed it could reasonably serve as a good competitor. The reason for using a simple curve instead of various other catheter shapes was to facilitate the comparison of the mechanical properties between the 2 systems. Third, the guided angiography system was performed with the same contrast injection protocol as the conventional angiography system, meaning that the protocol

was not optimized for the guided angiography system. Because the aneurysm diagnosis and decision-making were not hindered by image quality, we decided to continue the original study protocol and allow the suboptimal image quality as a limitation of this study. Further investigation is required to determine optimal injection protocols tailored for the guided angiography system.

## CONCLUSIONS

This study provides evidence that the guided angiography system is a safe, effective, and convenient alternative to the conventional angiography system for diagnostic cerebral conventional angiography. There were no significant differences in the safety profile, and no serious AEs were observed between the 2 groups. Further clinical experience is needed to fully evaluate the safety of the guided angiography system. Additionally, optimizing the contrast injection protocol may help improve the image quality.

## ACKNOWLEDGMENTS

We would like to thank Sang Ik Park, Hye Hyeon Moon, and So Jeong Lee, who helped us in supporting this study. The statistics department at our institution kindly provided statistical advice for this article.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Alakbarzade V, Pereira AC. **Cerebral catheter angiography and its complications.** *Pract Neurol* 2018;18:393–98 CrossRef Medline
2. Dawkins AA, Evans AL, Wattam J, et al. **Complications of cerebral angiography: a prospective analysis of 2,924 consecutive procedures.** *Neuroradiology* 2007;49:753–59 CrossRef Medline
3. Kaufmann TJ, Huston J 3rd, Mandrekar JN, et al. **Complications of diagnostic cerebral angiography: evaluation of 19,826 consecutive patients.** *Radiology* 2007;243:812–19 CrossRef Medline
4. Ahn SH, Prince EA, Dubel GJ. **Basic neuroangiography: review of technique and perioperative patient care.** *Semin Intervent Radiology* 2013;30:225–33 CrossRef Medline
5. Morris P. *Practical Neuroangiography.* Lippincott Williams & Wilkins; 2007
6. Bendzus M, Koltzenburg M, Bartsch AJ, et al. **Heparin and air filters reduce embolic events caused by intra-arterial cerebral angiography: a prospective, randomized trial.** *Circulation* 2004;110:2210–15 CrossRef Medline
7. Chuah KC, Stuckey SL, Berman IG. **Silent embolism in diagnostic cerebral angiography: detection with diffusion-weighted imaging.** *Australas Radiol* 2004;48:133–38 CrossRef Medline
8. Gupta R, Vora N, Thomas A, et al. **Symptomatic cerebral air embolism during neuro-angiographic procedures: incidence and problem avoidance.** *Neurocrit Care* 2007;7:241–46 CrossRef Medline
9. Mamourian AC, Weglarz M, Dunn J, et al. **Injection of air bubbles during flushing of angiocatheters: an in vitro trial of conventional hardware and techniques.** *AJNR Am J Neuroradiol* 2001;22:709–12 Medline
10. Kwon B, Lee KB, Yoon JT, et al. **Safety and effectiveness of the novel catheter 3.0 system for diagnostic cerebral angiography: a pilot study.** *Neurointervention* 2022;17:143–51 CrossRef Medline
11. Kwon B, Song Y, Hwang SM, et al. **Injection of contrast media using a large-bore angiography catheter with a guidewire in place: physical factors influencing injection pressure in cerebral angiography.** *Interv Neuroradiol* 2021;27:558–65 CrossRef Medline
12. Thiex R, Norbash AM, Frerichs KU. **The safety of dedicated-team catheter-based diagnostic cerebral angiography in the era of advanced noninvasive imaging.** *AJNR Am J Neuroradiol* 2010;31:230–34 CrossRef Medline
13. Cloft HJ, Jensen ME, Kallmes DF, et al. **Arterial dissections complicating cerebral angiography and cerebrovascular interventions.** *AJNR Am J Neuroradiol* 2000;21:541–45 Medline

# Reentry Technique for Rescue Recanalization of Carotid Near-Total Occlusion after Subintimal Penetration

 Dang Khoi Tran,  Huu-Thao Nguyen,  Chih-Wei Huang,  Kai-Chen Chung, and  Yuang Seng Tsuei

## ABSTRACT

**C**arotid stent placement is an effective alternative for stroke prevention in patients with carotid stenosis. However, endovascular recanalization of near-total carotid occlusion remains challenging, with subintimal dissection sometimes unavoidable. We present the case of an adult in their 70s with symptomatic carotid stenosis. During plaque crossing with a microwire, unintentional subintimal dissection occurred. The reentry technique, using a specialized reentry catheter, was performed, followed by carotid stent placement, achieving successful recanalization. The patient made a good clinical recovery. In selected cases, subintimal recanalization of carotid occlusion using the rescue re-entry technique can be performed safely after entry into the false lumen.<sup>1-5</sup>

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Limbucci N, Nappini S, Rosi A, et al. **Intentional subintimal carotid stenting of internal carotid dissection in a patient with acute ischaemic stroke.** *Interv Neuroradiol* 2017;23:69–72 CrossRef Medline
2. Lin MS, Lin LC, Li HY, et al. **Procedural safety and potential vascular complication of endovascular recanalization for chronic cervical internal carotid artery occlusion.** *Circ Cardiovasc Interv* 2008;1:119–25 CrossRef Medline
3. Morino Y, Kimura T, Hayashi Y et al; J-CTO Registry Investigators. **In-hospital outcomes of contemporary percutaneous coronary intervention in patients with chronic total occlusion insights from the J-CTO Registry (Multicenter CTO Registry in Japan).** *JACC Cardiovasc Interv* 2010;3:143–51 CrossRef Medline
4. Müller MD, Lyrer P, Brown MM, et al. **Carotid artery stenting versus endarterectomy for treatment of carotid artery stenosis.** *Cochrane Database Syst Rev* 2020;2:Cd000515 CrossRef Medline
5. Schneider PA, Caps MT, Nelken N. **Re-entry into the true lumen from the subintimal space.** *J Vasc Surg* 2013;58:529–34 CrossRef Medline

Received February 20, 2025; accepted after revision April 16.

From the Neurosurgery Department (D.K.T.), University Medical Center Ho Chi Minh City, Vietnam; Vascular Department (H.-T.N.), Cho Ray Hospital, Ho Chi Minh City, Vietnam; Neurosurgery Department (C.-W.H., K.-C.C., Y.S.T.), Taichung Veterans General Hospital, Taiwan; and National Chung Hsing University Taichung (Y.S.T), Taichung, Taiwan.

Please address correspondence to Yuang Seng Tsuei, MD, Neurosurgery Department, Taichung Veterans General Hospital, 1650 Taiwan Boulevard Section 4, Taichung, Taiwan 40705, ROC; e-mail: [astrocytoma2001@yahoo.com.tw](mailto:astrocytoma2001@yahoo.com.tw)

<http://dx.doi.org/10.3174/ajnr.A8860>

# Improving the Robustness of Deep Learning Models in Predicting Hematoma Expansion from Admission Head CT

Anh T. Tran,  Gaby Abou Karam, Dorin Zeevi,  Adnan I. Qureshi,  Ajay Malhotra,  Shahram Majidi, Santosh B. Murthy, Soojin Park, Despina Kontos, Guido J. Falcone,  Kevin N. Sheth, and  Seyedmehdi Payabvash



## ABSTRACT

**BACKGROUND AND PURPOSE:** Robustness against input data perturbations is essential for deploying deep learning models in clinical practice. Adversarial attacks involve subtle, voxel-level manipulations of scans to increase deep learning models' prediction errors. Testing deep learning model performance on examples of adversarial images provides a measure of robustness, and including adversarial images in the training set can improve the model's robustness. In this study, we examined adversarial training and input modifications to improve the robustness of deep learning models in predicting hematoma expansion (HE) from admission head CTs of patients with acute intracerebral hemorrhage (ICH).

**MATERIALS AND METHODS:** We used a multicenter cohort of  $n = 890$  patients for cross-validation/training, and a cohort of  $n = 684$  consecutive patients with ICH from 2 stroke centers for independent validation. Fast gradient sign method (FGSM) and projected gradient descent (PGD) adversarial attacks were applied for training and testing. We developed and tested 4 different models to predict  $\geq 3$  mL,  $\geq 6$  mL,  $\geq 9$  mL, and  $\geq 12$  mL HE in an independent validation cohort applying receiver operating characteristics area under the curve (AUC). We examined varying mixtures of adversarial and nonperturbed (clean) scans for training as well as including additional input from the hyperparameter-free Otsu multithreshold segmentation for model.

**RESULTS:** When deep learning models trained solely on clean scans were tested with PGD and FGSM adversarial images, the average HE prediction AUC decreased from 0.8 to 0.67 and 0.71, respectively. Overall, the best performing strategy to improve model robustness was training with 5:3 mix of clean and PGD adversarial scans and addition of Otsu multithreshold segmentation to model input, increasing the average AUC to 0.77 against both PGD and FGSM adversarial attacks. Adversarial training with FGSM improved robustness against similar type attack but offered limited cross-attack robustness against PGD-type images.

**CONCLUSIONS:** Adversarial training and inclusion of threshold-based segmentation as an additional input can improve deep learning model robustness in prediction of HE from admission head CTs in acute ICH.

**ABBREVIATIONS:** ATACH-2 = Antihypertensive Treatment of Acute Cerebral Hemorrhage; AUC = area under the curve; CNN = convolutional neural network; Dice = Dice coefficient; FGSM = fast gradient sign method; HD = Hausdorff distance; HE = hematoma expansion; ICH = intracerebral hemorrhage; PGD = projected gradient descent; ROC = receiver operating characteristic; VS = volume similarity

In addition to prediction accuracy, several factors determine the trustworthiness of deep learning models in health care, including robustness, generalizability, interpretability, fairness, and security.<sup>1</sup> Robustness refers to the model's ability to maintain performance despite noise or perturbations in input

data.<sup>2</sup> In medical image analysis, a robust deep learning model can accurately classify or segment clinical scans even in the presence of noise, corrupted voxels, or blurring. Main strategies to improve the robustness of deep learning models include adversarial training,<sup>2</sup> preprocessing, and postprocessing techniques such as feature squeezing,<sup>3</sup> and model designs that can detect perturbed image inputs (ie, adversarial attacks).<sup>4</sup>

Adversarial images are generated by intentional manipulation of original scans to include subtle, voxel-level perturbations

Received November 19, 2024; accepted after revision January 7, 2025.

From the Departments of Radiology (A.T.T., D.Z., S.P.), and Neurology (S.P., D.K.), New York-Presbyterian/Columbia University Irving Medical Center, Columbia University, New York, New York; Departments of Radiology and Biomedical Imaging (A.T.T., G.A.K., D.Z., A.M.), Neurology (G.J.F. K.N.S.), and Center for Brain and Mind Health (G.J.F., K.N.S.), Yale School of Medicine, New Haven, Connecticut; Zeenat Qureshi Stroke Institute and Department of Neurology (A.I.Q.), University of Missouri, Columbia, Missouri; Department of Neurosurgery, Icahn School of Medicine at Mount Sinai (S.M.), Mount Sinai Hospital, New York, New York; and Department of Neurology, Weill Cornell Medical College (S.B.M.), Cornell University, New York.

This work was supported by Doris Duke Charitable Foundation (2020097), NIH (K23NS118056), and NVIDIA Applied Research Accelerator Program.

Please address correspondence to Seyedmehdi Payabvash, MD, Center for Innovation in Imaging Biomarkers and Integrated Diagnostics, Department of Radiology, Columbia University, 530 West 166<sup>th</sup> St, 5<sup>th</sup> Floor, New York, NY 10032, e-mail: Sam.payabvash@gmail.com; @SamPayabvash; @Columbialmaging

 Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8650>

## SUMMARY

**PREVIOUS LITERATURE:** The success of deep learning models in medical imaging relies on their generalizability and stability of predictions. However, small, imperceptible voxel-level perturbations (adversarial attacks) can mislead models, reducing their accuracy and raising concerns about their clinical reliability. Prior studies suggest that adversarial training (ie, incorporating perturbed images in training set) can improve deep learning model robustness; however, optimal strategies for increasing the robustness of HE prediction remain unclear. Ensuring robustness in HE prediction models from admission head CTs is critical for their clinical use in guiding targeted treatment of at-risk patients with hemorrhagic stroke.

**KEY FINDINGS:** Training with a mixture of perturbed and clean (nonperturbed) noncontrast head CT scans can effectively improve the robustness of deep learning models for hematoma segmentation and HE prediction (classification). Additionally, incorporating input from hyperparameter-free Otsu threshold-based segmentation of head CTs can further increase the robustness of these models.

**KNOWLEDGE ADVANCEMENT:** We reported the optimal adversarial training strategy and the benefits of adding threshold-based Otsu segmentation to improve the robustness of hematoma segmentation and HE prediction deep learning models. These models can guide targeted therapies in hemorrhagic stroke, and our proposed methodology can be extended to development of other robust deep learning models.

designed to maximize prediction errors of deep learning models.<sup>1</sup> These voxel-level changes are often imperceptible to human eyes and different from introducing random noise into images because adversarial images are purposely crafted to challenge a model's prediction, and are more effective in evaluating a model's robustness against input perturbations.<sup>1</sup> The vulnerability of deep learning models to adversarial images (attacks) raises concerns about their robustness and trustworthiness in real-world clinical practice.<sup>5</sup>

Adversarial training refers to including adversarial images in the training set and is one of the most computationally efficient strategies for enhancing the robustness of deep learning models.<sup>2</sup> However, training models exclusively with adversarial images may reduce their accuracy on unperturbed inputs.<sup>5,6</sup> In this study, we systematically applied and optimized adversarial training to improve the robustness of deep learning models for predicting hematoma expansion (HE) from admission noncontrast head CT scans of patients with acute intracerebral hemorrhage (ICH). To improve model robustness further, we also incorporated an automated thresholding step during image processing. HE affects nearly one-third of patients with ICH within 6 hours of admission and is an independent predictor of neurologic deterioration, disability, and mortality.<sup>7</sup> Predicting HE is clinically valuable because it can guide targeted antiexpansion and hemostatic therapies in ICH.<sup>8,9</sup> Recent studies have shown that radiomic features of the hematoma and deep learning models can predict HE from admission noncontrast head CTs with greater accuracy than visual markers determined by expert reviewers, such as the blend sign and swirl sign.<sup>10,11</sup> Improving the robustness of HE prediction models can make them more suitable for the real-world clinical practice.

## MATERIALS AND METHODS

### Patient Ascertainment

We used the admission and follow-up head CT scans from the Antihypertensive Treatment of Acute Cerebral Hemorrhage (ATACH-2) multicenter randomized trial for training and cross-validation.<sup>12</sup> Briefly, ATACH-2 evaluated intensive blood

pressure reduction in patients presenting with spontaneous ICH and at least 1 systolic blood pressure >180 mm Hg but found no treatment benefit.<sup>12</sup> The trial included 110 sites in the United States, Germany, Japan, China, Taiwan, and South Korea.<sup>12</sup> For external validation, we tested the models in an independent cohort of consecutive patients presenting with spontaneous ICH to Yale Health Stroke Centers (Yale New Haven Hospital at York Street and Saint Raphael Campuses) from January 2015 to December 2023. Patients were included if they had baseline and follow-up noncontrast head CT scans within 6 and 36 hours after the onset, respectively. Subjects with metal/streak artifacts or surgical interventions affecting hematoma lesions on either baseline or follow-up scan, precluding accurate segmentation of hematoma, were excluded from the analysis. The institutional review boards of participating centers in the ATACH-2 clinical trial approved follow-up analysis of data. Our retrospective study of patients with ICH at Yale received separate institutional review board approval.

### Generation of Ground Truth Labels for HE

We manually segmented hematoma lesions on baseline and follow-up head CTs, as described previously.<sup>10,11,13</sup> Segmentations were performed by trained research associates and then reviewed/revised by a board-certified neuroradiologist with more than 10 years of experience. The intra- and interrater reliability of segmentations were determined in a subset of scans by using intraclass correlation, ranging from 0.92 to 0.94.<sup>10,11,13</sup> We trained and validated separate models for prediction of  $\geq 3$  mL,  $\geq 6$  mL,  $\geq 9$  mL, and  $\geq 12$  mL HE from baseline to follow-up scan as binary classifications.<sup>14</sup> These HE thresholds were described originally to predict poor outcomes with increasing specificity and positive predictive values in patients with ICH.<sup>14</sup>

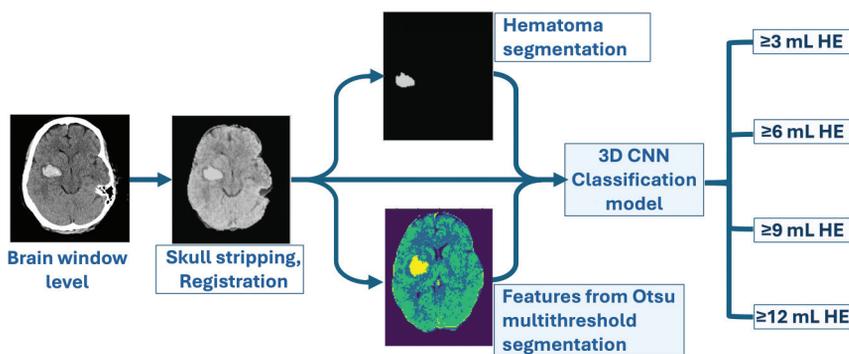
### Backbone of the HE Prediction Model

Our image analysis pipeline included preprocessing steps for skull removal,<sup>11</sup> adjustment to brain window-level, and coregistration of brain to isotropic 3D template,<sup>15</sup> as detailed in the Supplemental Data. Our prior experience showed that a dual input from CT

slices and hematoma mask will improve the classification performance for HE prediction.<sup>11</sup> For the backbone of hematoma segmentation model, we used nnUNET,<sup>16</sup> a versatile U-Net-shaped convolutional neural network (CNN) with self-configuration capabilities.<sup>16</sup> For classification, we modified the DenseNet121 3D CNN for prediction of HE. We developed and tested 4 different models for prediction of  $\geq 3$  mL,  $\geq 6$  mL,  $\geq 9$  mL, and  $\geq 12$  mL HE. The overall model structure is depicted in Fig 1 and further described below.

### Creating Adversarial Images

In this study, we applied fast gradient sign method (FGSM)<sup>17</sup> and projected gradient descent (PGD)<sup>18</sup> to create adversarial images from brain CT. These methods are designed to maximize classification error of the deep learning model while minimizing the difference between the adversarial and original images. All adversarial image generation methods are bounded under a predefined perturbation size  $\epsilon$ , which represents the maximum change added to each pixel/voxel value, after normalizing their intensities to 0–1 range. Details of the 2 methods are included in Supplemental Data, and examples are depicted in Fig 2.



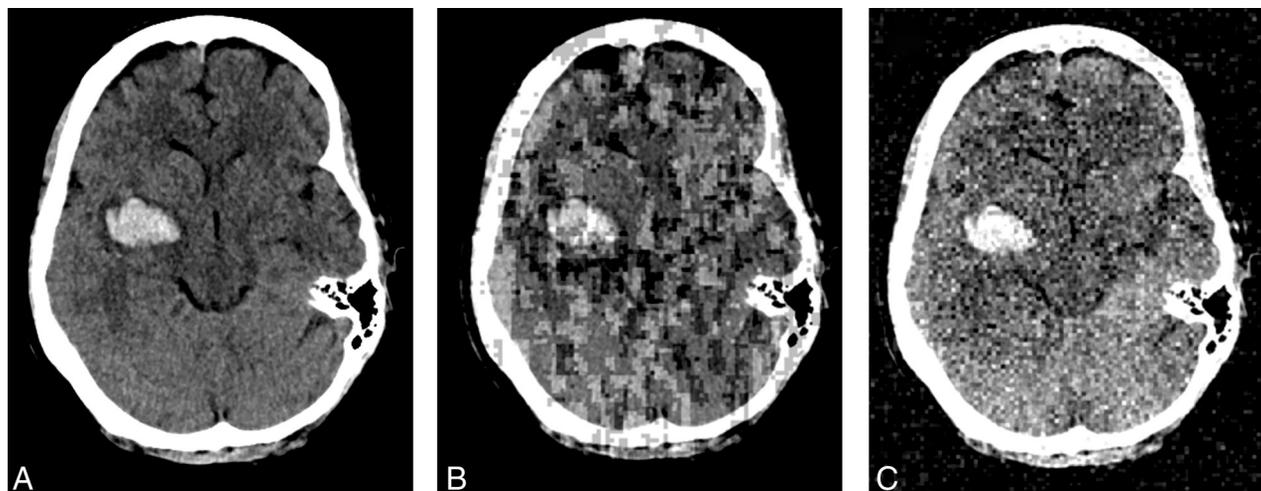
**FIG 1.** Our final model included preprocessing of head CTs with adjusting the images to brain window level (80:40 Hounsfield units), skull removal, and registration to a similar size  $128 \times 128 \times 128$  template. The model included (3, 128, 128, 128) inputs from skull-stripped head CT slices, automated hematoma segmentation masks, and Otsu multithreshold segmentation. We trained and tested 4 separate model for prediction of  $\geq 3$  mL,  $\geq 6$  mL,  $\geq 9$  mL, and  $\geq 12$  mL HE.

### Adversarial Training to Improve Model Robustness

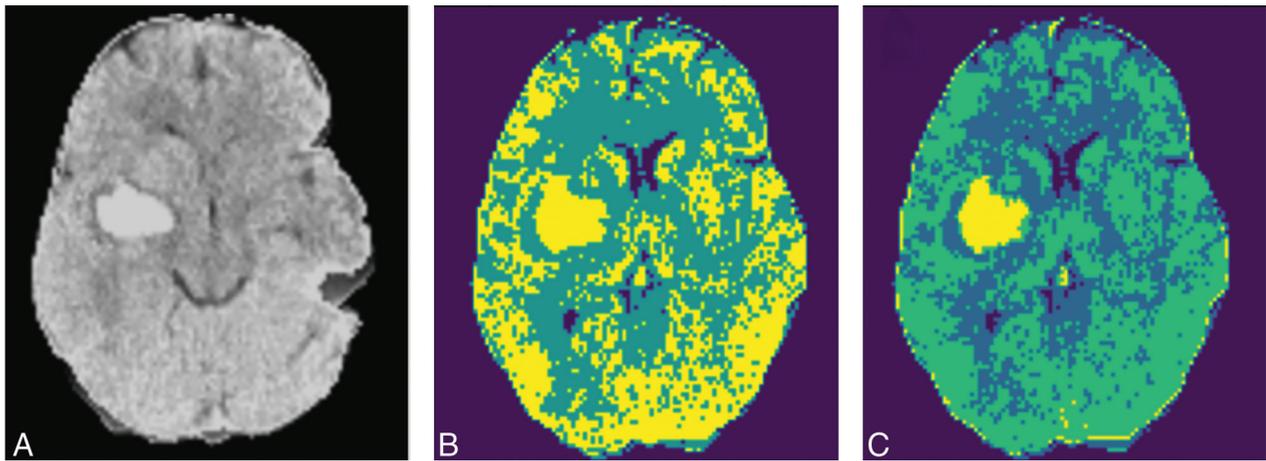
Inclusion of adversarial images in a training data set can improve the resilience of deep learning models to adversarial attacks by encouraging the models to learn more robust and meaningful features, rather than overfitting to cohort-specific patterns of the training data set. However, training of models exclusively with adversarial images can reduce their accuracy for unperturbed input. By exposing the model to both adversarial and original (clean) images during training, the deep learning model will learn to classify both image types more accurately compared with being trained only on 1 image type. In this study, we trained the model with varying mixtures of clean-to-adversarial images (from 5:5 to 5:1) to identify the optimal strategy for maximum accuracy. Clean images included original scans, and basic augmentations such as resize, rotate, and flip without adversarial perturbations. Adversarial images included perturbed variations of original scans and basic augmentations.

### Otsu Multithreshold Segmentation to Improve Robustness

The Otsu thresholding method is a popular technique in image segmentation due to its simplicity and efficiency.<sup>19</sup> The algorithm automatically separates voxels of input images into a predefined number of classes based on their signal intensity by maximizing the between-class variance.<sup>19</sup> Because such multiclass segmentations maintain the overall spatial consistency of the image, their outputs are more resilient to subtle, inconsistent changes introduced by adversarial attacks. Multiclass segmentation of input images can improve the robustness of final deep learning models and counter adversarial attacks.<sup>20</sup> The task of distinguishing between multiple classes at a pixel/voxel level forces the model to become more



**FIG 2.** Examples of (A) an original head CT slice, and (B) adversarial images after applying FGSM,<sup>17</sup> and (C) PGD,<sup>18</sup> with  $\epsilon = 0.1$  perturbation size.



**FIG 3.** Example of (A) the skull-stripped and registered brain CT, with (B) 3-level and (C) 4-level Otsu thresholding multiclass segmentation. The results of 4-level thresholding had better overlap with hematoma lesions.

discriminative. This heightened class-wise differentiation reduces the impact of adversarial attacks, which tend to confuse the deep learning models by blending features of different classes. In addition, multithreshold segmentation often requires capturing overlapping features between different classes. This redundancy can act as a form of defense because perturbations that affect 1 part of the image may not be sufficient to fool the entire segmentation map. The model can use redundancy to cross-verify predictions, making it harder for an adversarial attack to succeed. We added Otsu multithreshold segmentation to both segmentation and classification pipelines. Thus, the inputs for the classification CNN were a concatenation of brain images, hematoma segmentation mask, and the features from Otsu thresholding multiclass segmentation. In our exploratory analysis, we found that 4-level thresholding has better delineation of hemorrhage in brain tissue based on average Dice results (Fig 3).

**Training and Validation of Hematoma Segmentation Model.** We used 5-fold cross-validation for training the hematoma segmentation model. The base model was the nnUNet Version 2.0, and we included Otsu multithreshold segmentation as an additional input to head CT slices. For training we applied  $\text{weight\_decay} = 3e-5$ ,  $\text{initial learning rate} = 0.001$ ,  $\text{num\_epochs} = 100$ , PolyLRScheduler, optimizer = SGD, and patch window size = (96, 96, 128). These hyperparameters were optimized through a grid search strategy. The model accuracy was determined by using Dice, Hausdorff distance (HD), and volume similarity (VS) as described in the Supplemental Data. The Supplemental Data represent an example of loss function and Dice diagram during the training process. Robustness of segmentation model was evaluated by mean Dice, HD, and VS in independent test cohorts without and with adversarial images of various perturbation sizes.

**Training and Validation of HE Prediction Models.** For training of HE prediction models, we used similar 5-fold cross-validation split as segmentation model, applying DenseNet121 from MONAI platform with  $\text{loss\_function} = \text{BCEWithLogitsLoss}()$ , optimizer = Adam(), scheduler = ReduceLROnPlateau, learning rate = 0.001,  $\text{weight\_decay} = 1e-4$ , augmentation = (RandFlip, RandZoom,

RandRotate),  $\text{batch\_size} = 10$ . These hyperparameters were optimized through a grid search strategy. The inputs (3, 128, 128, 128) of our final model were skull-stripped brain CT slices, automated hematoma segmentation masks, and Otsu 4-class thresholding segmentation. We developed and tested separate models for prediction of  $\geq 3$  mL,  $\geq 6$  mL,  $\geq 9$  mL, and  $\geq 12$  mL HE. We tested model prediction performance by using the area under the curve (AUC) of receiver operating characteristics (ROC) analysis. Supplemental Data represent an example of loss function and AUC changes during training process. We compared the effectiveness of the proposed strategies in improving model robustness by testing their prediction AUC in clean (nonperturbed) and perturbed images.

## RESULTS

### Patient Characteristics

Of 1000 patients enrolled in the multicentric ATACH-2 trial,  $n = 890$  were included in cross-validation/training cohort, 11 were excluded due to CT and surgical hardware artifacts, 59 due to missing follow-up head CT scans, and 40 due to missing clinical information. From 940 patients with ICH in the Yale stroke registry,  $n = 684$  were included in the independent validation cohort, 10 excluded due to CT and surgical hardware artifacts, 97 due to missing baseline or follow-up head CT scans, and 149 due to missing clinical information. Details of patient characteristics and CT scan information are listed in the Table. Overall, patients in the independent validation cohort were more likely to be women, be older, and have larger baseline and follow-up hematoma volumes (all  $P < .001$ ).

### Hematoma Segmentation Model

The Supplemental Data provide details of hematoma segmentation performance across different model structures and with adversarial training. Using clean (nonperturbed) CT scans for training with the baseline nnUNET model, we achieved an average Dice score of  $0.91 \pm 0.14$  for hematoma segmentation in clean CT scans. However, the Dice of the model trained on clean CT images decreased to  $0.24 \pm 0.29$ , and  $0.71 \pm 0.19$  when tested on FGSM and PGD adversarial images ( $\epsilon = 0.1$ ). By adding Otsu

multilevel thresholding to model input, the average Dice for segmentation on FGSM and PGD adversarial images improved to

$0.51 \pm 0.37$  and  $0.81 \pm 0.17$ , respectively, while maintaining a similar average Dice score in clean CTs. Further improvements in

adversarial performance were achieved by training with a 1:1 mixture of clean and adversarial images by using segmentation model with additional input from Otsu. Training with a mixture of clean scans and FGSM adversarial images yielded average Dice scores of  $0.83 \pm 0.20$  and  $0.73 \pm 0.19$  when tested on FGSM and PGD adversarial images, respectively. Training with a mixture of clean scans and PGD adversarial images resulted in average Dice scores of  $0.83 \pm 0.20$  and  $0.85 \pm 0.24$  when tested on FGSM and PGD adversarial images, respectively. Segmentation performance on clean CT scans was maintained, with an average Dice score of 0.91.

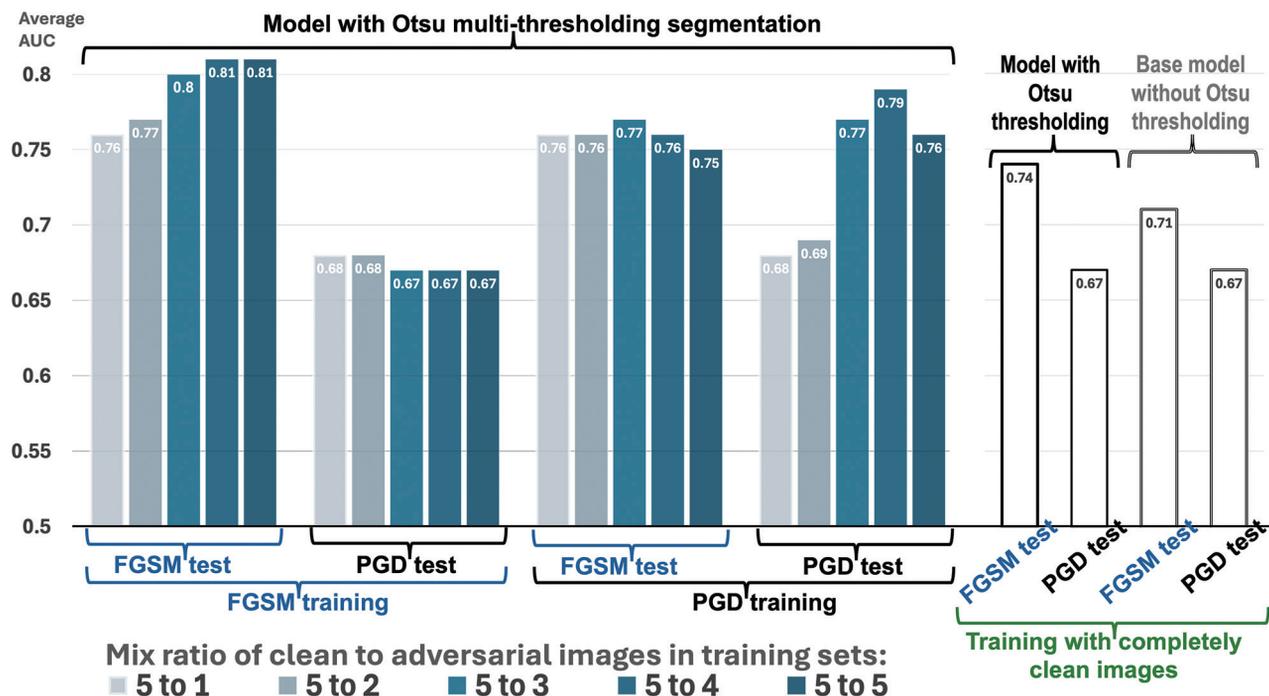
**The demographic and clinical characteristics of patients in training/cross-validation versus independent test cohort**

Column A	Training/Cross-Validation from ATACH-2 (n=890)	Independent Test (n=684)	P Value
Sex, man	543 (60.8%)	358 (54.8%)	<.001
Age [years]	62.15 ± 13.1	69.7 ± 14.3	<.001
Hypertension	706 (79.3%)	583 (85.2%)	.17
Diabetes	172 (19.3%)	181 (26.5%)	.28
Hyperlipidemia	221 (24.8%)	332 (48.5%)	.81
Atrial fibrillation	30 (3.4%)	150 (21.9%)	.17
NIHSS score at baseline			.86
0–4	148 (16.57%)	243 (35.57%)	
5–9	235 (26.31%)	119 (17.40%)	
10–14	241 (26.98%)	81 (11.85%)	
15–19	161 (18.02%)	92 (13.46%)	
20–25	71 (7.95%)	52 (7.60%)	
>25	37 (4.17%)	30 (4.39%)	
Unknown		66 (9.73%)	
Baseline hematoma volume [mL]	12.9 ± 12.6	18.7 ± 20.6	<.001
Follow-up hematoma volume [mL]	15.6 ± 16.6	22.9 ± 25.8	<.001
CT voxel spacing [mm]	[0.46 ± 0.09, 0.47 ± 0.09]	[0.46 ± 0.04, 0.46 ± 0.04]	
Slice thickness [mm]	5.20 ± 1.86	4.81 ± 0.69	
Min axial [n × n]	[418 × 418]	[472 × 472]	
Max axial [n × n]	[512 × 734]	[1024 × 1024]	
Number of slices	31.2 ± 18.0	35.0 ± 11.2	

Data are presented as number (percentage) or mean ± standard deviation. Nominal variables are compared by using  $\chi^2$  and continuous variables are compared by using t-test.

**Improving Robustness of HE Prediction Models**

The Supplemental Data summarize the performance of different HE prediction models with and without adversarial training. Figure 4 presents the performance of various mixtures of clean and adversarial images in training when tested on FGSM and PGD adversarial



**FIG 4.** Different mix ratios of FGSM and PGD adversarial images with clean (nonperturbed) scans were used to train HE prediction models (with model input including brain CT slices, automatically segmented hematoma masks, and Otsu multilevel thresholding). The average AUC ROC for  $\geq 3$  mL,  $\geq 6$  mL,  $\geq 9$  mL, and  $\geq 12$  mL HE prediction models are shown, when tested against adversarial attacks with  $\epsilon = 0.1$  perturbation size. PGD attacks were more effective in reducing the AUCs than FGSM attack. On the other hand, PGD training also improved model performance against FGSM attacks. Notably, addition of Otsu-thresholding in the prediction model slightly improved model performance against FGSM attacks even in the absence of any adversarial training.

images with the highest perturbation level applied in our study ( $\epsilon = 0.1$ ). The average AUCs of HE prediction models by using inputs from brain CT slices and hematoma segmentations were 0.8 on clean scans, dropping to 0.67 and 0.71 when tested on FGSM and PGD adversarial images ( $\epsilon = 0.1$ ). Notably, the hematoma masks were generated via automated segmentation by a model optimized with adversarial training. Including Otsu multithreshold segmentation in the prediction model improved the average AUC on FGSM adversarial images ( $\epsilon = 0.1$ ) to 0.74 (Supplemental Data). Overall, the PGD adversarial attacks with  $\epsilon$  values of 0.01 and 0.1 reduced model performance more than FGSM attacks. Adversarial training to each specific attack type improved model robustness against similar attack, ie, a mixture of FGSM and clean images in training improved models' AUC when tested against FGSM attacks more effectively than against PGD, and vice versa. However, there was a trend indicating that PGD adversarial training also provided some improvement against FGSM attacks, whereas FGSM adversarial training offered limited performance gains against PGD attacks (Fig 4). Overall, a 5:3 mixture of clean and PGD adversarial images yielded the highest average AUC in HE prediction against both PGD and FGSM adversarial attacks. Any combination of clean and adversarial images during training preserved the baseline average AUC of 0.8 in HE prediction when tested on clean scans.

## DISCUSSION

Successful application of deep learning models in clinical practice depends on their generalizability and stability against perturbations in input data. Several recent studies have reported the vulnerability of medical imaging classification and segmentation models to adversarial attacks or small perturbations in input data.<sup>5,21,22</sup> We found that subtle voxel-level perturbations negatively impact the performance of both ICH segmentation and HE prediction models. Then, we showed that adversarial training with a mix of clean and perturbed scans can improve the robustness of hematoma segmentation and HE prediction models while maintaining accuracy in clean data sets. We also found cross-attack robustness of hematoma segmentation and HE prediction model against FGSM attacks after training on PGD adversarial images. From a computational efficiency perspective, our findings suggest that training on a mix of clean and a subset of adversarial images (eg, a 5:3 ratio) can sufficiently enhance model robustness against various types of input perturbations. In addition, we showed the benefit of including Otsu multithreshold segmentation as an additional input to improve both segmentation and classification robustness. The Otsu algorithm provides a simple, efficient, and parameter-free method by automatically selecting the optimal threshold to maximize between-class variance in the segmented image.<sup>19,20</sup> Overall, the key takeaway from our findings is 2-fold: first, to include a subset of adversarial images as part of data augmentation during model training; and second, to incorporate hyperparameter-free, threshold-based segmentation as an additional input for the model. These strategies may not increase model accuracy on clean (nonperturbed) test cohorts, but they can improve the model's robustness and stability in handling noisy and distorted images.

Identifying patients with ICH at risk of HE is crucial for guiding hemorrhagic stroke treatment. As an independent and potentially modifiable risk factor for poor outcomes and mortality, HE has been the target of antiexpansion and hemostatic therapies.<sup>12,23-25</sup> Recent clinical trials showed the benefits of intensive blood pressure reduction as part of a critical care bundle or in the prehospital setting.<sup>8,9</sup> Post hoc analyses of prior clinical trials also showed improved outcomes and reduced HE after ultra-early blood pressure reduction, despite negative overall trial results.<sup>26-28</sup> This highlights the need for reliable identification of patients with ICH at risk for HE, which can expand the eligibility of patients to receive intensive therapies, and expedite potential interhospital transfer decisions in acute settings. Recent studies have shown that radiomics and deep learning models can provide more reliable, automated prediction of HE.<sup>10,11,29-31</sup> However, for clinical practice use, these models require extensive external validation to confirm generalizability and robustness. Our findings provide a framework for improving the robustness of HE prediction models, paving the way for their potential clinical application.

Clean data might not fully represent the diverse range of inputs a deep learning model could encounter in real-world scenarios. Robustness examines the potential failure of models when exposed to perturbed data points. Adversarial images represent potential inputs that are sparsely represented in the clean data set, making accurate predictions challenging for the model. Notably, ambiguous and outlier images cannot be modeled by simply adding noise, as adversarial attacks are specifically crafted to induce prediction failures.<sup>1</sup> This makes them better suited for evaluating model behavior under extreme inputs. Adversarial training acts as a form of regularization, helping to prevent overfitting and improve model stability. By exposing the model to both clean and adversarial examples—as part of data augmentation during training—the model learns to balance the trade-off between accuracy on clean data and robustness to adversarial perturbations, resulting in more stable predictions.<sup>1,2,21</sup> As also shown in our study, features learned during adversarial training are often transferable across different types of adversarial attacks. This suggests that even if the attacks encountered during testing differ from those encountered in training, the model may still leverage the knowledge gained from training to defend against other forms of adversarial attacks.

Prior studies suggest that adversarial training can improve deep learning model accuracy, especially in underrepresented examples. Liu et al<sup>32</sup> reported that the addition of adversarial images and adversarial synthetic nodules to the training data improved model robustness in detecting underrepresented lung nodules and resistance to noise perturbations in low-dose CT scans. Vatian et al<sup>33</sup> showed that the addition of adversarial images can improve classification accuracy of chest CTs with lung nodules and brain MRIs with gliomas. Hu et al<sup>34</sup> found that adversarial training improves the generalizability and accuracy of deep learning models in the classification of prostate cancer MRIs. However, an analysis of a large x-ray data set ( $n=22,433$ ) suggests no improvement in deep learning model performance with adversarial training when trained with a large sample size.<sup>35</sup> In our study, there was no significant change in model performance on the clean test cohort with training on a mix of clean and adversarial images. However, the primary goal of adversarial training is to maintain model

performance when exposed to examples not included in the existing data set, as evidenced by the robustness of adversarially trained models against attacks in both segmentation and classification tasks.

We found that subtle voxel-level perturbations (eg,  $\epsilon = 0.001$  or  $0.01$ ) had a limited impact on the accuracy of hematoma segmentation and HE prediction models, whereas moderate perturbations ( $\epsilon = 0.1$ ) significantly diminished the performance of both segmentation and prediction models. This is consistent with prior studies showing a gradual decrease in the performance of medical image analysis models when  $\epsilon > 0.01$ .<sup>5,21,22</sup> Notably, perturbations with  $\epsilon > 0.1$  typically become visually conspicuous, rendering such images beyond the scope of model robustness improvement.<sup>5,21,22</sup> In addition, the impact of FGSM attacks on segmentation model performance was greater than that of PGD. This is likely because PGD, being an iterative method, refines its perturbations to maximize misclassification over multiple steps, resulting in more localized perturbations than the broader signal changes introduced by FGSM. Thus, when attacking segmentation models, PGD's local adjustments can retain some structural integrity, leading to better Dice scores compared with FGSM attack.

The Otsu method is a hyperparameter-free image segmentation technique that automatically identifies the optimal threshold(s) to maximize between-class variance of voxels in segmentation classes.<sup>19,20</sup> For the Otsu algorithm, the number of threshold levels must be predefined, and our experiments showed that a 4-level threshold on skull-stripped brain CT slices achieves the best overlap with hematoma segmentation masks. We found that adding Otsu multithreshold segmentation results as feature extraction in the model input improved the robustness of both hematoma segmentation and HE prediction models. This is likely because clustering different intensity levels within the image enhances the features of underlying pathologies (like hematoma) and reduces the influence of noise. With multilevel thresholding, the impact of minor perturbations is diminished, as the image is segmented into broader intensity-based regions. Thus, Otsu's method can act as a denoising layer, reducing the model's sensitivity to random noise and adversarial changes.<sup>36</sup>

There are inherent differences in computational efficiency of adversarial versus standard training process. Standard training is computationally more efficient, without generating additional adversarial images as part of data augmentation. This results in faster training times and reduced hardware requirements, making standard training ideal for resource-constrained settings. By contrast, adversarial training requires generating perturbed examples through techniques such as FGSM or PGD. These methods increase the computational load due to additional forward and backward passes, leading to significantly longer training times and the need for high-performance servers. On the other hand, inference efficiency typically remains similar for both training approaches, but the benefits of adversarial training may extend to improved model confidence and performance in challenging scenarios. For settings with limited computational resources, hybrid strategies, such as introducing adversarial examples gradually through cumulative learning, can help balance robustness and training efficiency. These approaches reduce the overall computational burden while still enhancing model robustness. In our analysis, adversarial training increased the training time by 3 times compared with standard

method, whereas testing of each patient took  $35 \pm 10$  seconds depending on the number of slices for either model type.

The strengths of our study include the use of a multicenter data set for training and an independent validation cohort. We also explored different mix ratios in adversarial training and demonstrated the benefits of multithreshold segmentation in enhancing the robustness of deep learning models. Aside from adversarial training, other methods to improve deep learning robustness include conventional data augmentation to expand training data sets, ensemble learning boosting or voting, regularization to prevent overfitting, feature squeezing in pre- or post-processing steps,<sup>3</sup> and identification of perturbed image inputs.<sup>4</sup> In addition, approaches such as transfer learning, domain adaptation, cross-validation, and ensemble learning can improve generalization and model adaptation across different data sets. In our study, aside from adversarial training, strategies such as data augmentation, regularization, dropouts, and normalization contributed to model robustness and generalizability. Future studies can combine these methods to further improve deep learning model robustness and generalizability, with the ultimate goal of maintaining model performance across different centers in clinical practice.

Our study also has several limitations. Our training cohort was limited by the inclusion and exclusion criteria of the ATACH-2 trial. The benefits of Otsu multilevel thresholding may also be limited to ICH, given the inherent contrast between hematoma and unaffected brain tissue, and may not generalize to other clinical scenarios. Other defense mechanisms against adversarial attacks, such as feature squeezing or adversarial sample detection, were not evaluated in this study. Although focusing on the brain window/level setting was the optimal choice for HE prediction based on our prior experience, other window/level settings may enable the extraction of additional CNN features. Exploring hybrid approaches or multi-window preprocessing in future iterations could further enhance the model's optimization for clinical applications. It should be noted that the primary goal of improving deep learning model robustness is to maintain performance stability against image perturbations caused by data heterogeneity and noise, whereas, addressing major metal or motion artifacts requires dedicated artifact reduction models.

## CONCLUSIONS

We found that hematoma segmentation and HE prediction models trained exclusively on nonperturbed (clean) head CT scans are vulnerable to voxel-intensity modifications from adversarial attacks. Our findings show that training with a mix of adversarial and clean data sets, along with incorporating threshold-based segmentation of brain CTs as additional input, can improve model robustness against adversarial attacks. Additionally, training on 1 type of adversarial images can provide cross-attack robustness against other types. Overall, our results suggest that including adversarial images as a subset of data augmentation instances in the training process, along with incorporating hyperparameter-free thresholding as an additional input, can improve the robustness of classification and segmentation deep learning models. Increased robustness of deep learning models will lead to higher stability when exposed to variations in input data, such as noise, artifacts, and visually imperceptible perturbations.

## REFERENCES

1. Paschali M, Conjeti S, Navarro F, et al. **Generalizability vs. robustness: adversarial examples for medical imaging.** *arXiv.org* 2018. <https://doi.org/10.48550/arXiv.1804.00504>. Accessed May 22, 2025
2. Apostolidis KD, Papakostas GA. **A survey on adversarial deep learning robustness in medical image analysis.** *Electronics* 2021;10:2132 CrossRef
3. Xu W, Evans D, Qi Y. **Feature squeezing: detecting adversarial examples in deep neural networks.** *arXiv.org* 2017. <https://doi.org/10.48550/arXiv.1704.01155>. Accessed May 22, 2025
4. Li X, Zhu D. **Robust Detection of Adversarial Attacks on Medical Images.** In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, Iowa. 2020:1154–58 CrossRef
5. Joel MZ, Umrao S, Chang E, et al. **Using adversarial images to assess the robustness of deep learning models trained on diagnostic images in oncology.** *JCO Clin Cancer Inform* 2022;6:e2100170 CrossRef Medline
6. Rodriguez D, Nayak T, Chen Y, et al. **On the role of deep learning model complexity in adversarial robustness for medical images.** *BMC Med Inform Decis Mak* 2022;22:160 CrossRef Medline
7. Abou Karam G, Chen MC, Zeevi D, et al. **Time-dependent changes in hematoma expansion rate after supratentorial intracerebral hemorrhage and its relationship with neurological deterioration and functional outcome.** *Diagnostics (Basel)* 2024;14:308 CrossRef Medline
8. Li G, Lin Y, Yang J, et al; INTERACT4 Investigators. **Intensive ambulance-delivered blood-pressure reduction in hyperacute stroke.** *N Engl J Med* 2024;390:1862–72 CrossRef Medline
9. Ma L, Hu X, Song L, et al; INTERACT3 Investigators. **The Third Intensive Care Bundle with Blood Pressure Reduction in Acute Cerebral Haemorrhage Trial (INTERACT3): an international, stepped wedge cluster randomised controlled trial.** *Lancet* 2023; 402:27–40 CrossRef Medline
10. Haider SP, Qureshi AI, Jain A, et al. **Radiomic markers of intracerebral hemorrhage expansion on non-contrast CT: independent validation and comparison with visual markers.** *Front Neurosci* 2023;17:1225342 CrossRef Medline
11. Tran AT, Zeevi T, Haider SP, et al. **Uncertainty-aware deep-learning model for prediction of supratentorial hematoma expansion from admission non-contrast head computed tomography scan.** *NPJ Digit Med* 2024;7:26 CrossRef Medline
12. Qureshi AI, Palesch YY, Barsan WG, et al; ATACH-2 Trial Investigators and the Neurological Emergency Treatment Trials Network. **Intensive blood-pressure lowering in patients with acute cerebral hemorrhage.** *N Engl J Med* 2016;375:1033–43 CrossRef Medline
13. Haider SP, Qureshi AI, Jain A, et al. **The coronal plane maximum diameter of deep intracerebral hemorrhage predicts functional outcome more accurately than hematoma volume.** *Int J Stroke* 2022; 17:777–84 CrossRef Medline
14. Dowlatshahi D, Demchuk AM, Flaherty ML, et al; VISTA Collaboration. **Defining hematoma expansion in intracerebral hemorrhage: relationship with patient outcomes.** *Neurology* 2011; 76:1238–44 CrossRef Medline
15. Rorden C, Bonilha L, Fridriksson J, et al. **Age-specific CT and MRI templates for spatial normalization.** *Neuroimage* 2012;61:957–65 CrossRef Medline
16. Isensee F, Jaeger PF, Kohl SAA, et al. **nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation.** *Nat Methods* 2021;18:203–11 CrossRef Medline
17. Goodfellow IJ, Shlens J, Szegedy C. **explaining and harnessing adversarial examples.** *arXiv.org* 2014. <https://arxiv.org/abs/1412.6572>. Accessed May 22, 2025
18. Madry M AA, Schmidt L, Tsipras D, et al. **Towards deep learning models resistant to adversarial attacks.** *arXiv.org* 2017. <https://arxiv.org/abs/1706.06083>. Accessed May 22, 2025
19. Otsu N. **A threshold selection method from gray-level histograms.** *IEEE Trans Syst Man Cybern* 1979;9:62–66 CrossRef
20. Jing Z, Tang B. **Improved image segmentation method based on Otsu thresholding and level set techniques.** *J Phys: Conf Ser* 2024; 2813:012017 CrossRef
21. Bortsova G, Gonzalez-Gonzalo C, Wetstein SC, et al. **Adversarial attack vulnerability of medical image analysis systems: unexplored factors.** *Med Image Anal* 2021;73:102141 CrossRef Medline
22. Muoka GW, Yi D, Ukwuoma CC, et al. **A comprehensive review and analysis of deep learning-based medical image adversarial attack and defense.** *Mathematics* 2023;11:4272 CrossRef
23. Gladstone DJ, Aviv RI, Demchuk AM, et al; SPOTLIGHT and STOP-IT Investigators and Coordinators. **Effect of recombinant activated coagulation factor vii on hemorrhage expansion among patients with spot sign-positive acute intracerebral hemorrhage: the SPOTLIGHT and STOP-IT Randomized Clinical Trials.** *JAMA Neurol* 2019;76:1493–501 CrossRef Medline
24. Meretoja A, Yassi N, Wu TY, et al. **Tranexamic acid in patients with intracerebral haemorrhage (STOP-AUST): a multicentre, randomised, placebo-controlled, phase 2 trial.** *Lancet Neurol* 2020;19: 980–87 CrossRef Medline
25. Naidech AM, Grotta J, Elm J, et al. **Recombinant factor VIIa for hemorrhagic stroke treatment at earliest possible time (FASTEST): protocol for a phase III, double-blind, randomized, placebo-controlled trial.** *Int J Stroke* 2022;17:806–09 CrossRef Medline
26. Leasure AC, Qureshi AI, Murthy SB, et al. **Intensive blood pressure reduction and perihematomal edema expansion in deep intracerebral hemorrhage.** *Stroke* 2019;50:2016–22 CrossRef Medline
27. Leasure AC, Qureshi AI, Murthy SB, et al. **Association of intensive blood pressure reduction with risk of hematoma expansion in patients with deep intracerebral hemorrhage.** *JAMA Neurol* 2019; 76:949–55 CrossRef Medline
28. Li Q, Warren AD, Qureshi AI, et al. **Ultra-early blood pressure reduction attenuates hematoma growth and improves outcome in intracerebral hemorrhage.** *Ann Neurol* 2020;88:388–95 CrossRef Medline
29. Teng L, Ren Q, Zhang P, et al. **Artificial intelligence can effectively predict early hematoma expansion of intracerebral hemorrhage analyzing noncontrast computed tomography image.** *Front Aging Neurosci* 2021;13:632138 CrossRef Medline
30. Chen Q, Zhu D, Liu J, et al. **Clinical-radiomics nomogram for risk estimation of early hematoma expansion after acute intracerebral hemorrhage.** *Acad Radiology* 2021;28:307–17 CrossRef Medline
31. Ma C, Zhang Y, Niyazi T, et al. **Radiomics for predicting hematoma expansion in patients with hypertensive intraparenchymal hematomas.** *Eur J Radiology* 2019;115:10–15 CrossRef Medline
32. Liu S, Setio AAA, Ghesu FC, et al. **No surprises: training robust lung nodule detection for low-dose CT scans by augmenting with adversarial attacks.** *IEEE Trans Med Imaging* 2021;40:335–45 CrossRef Medline
33. Vatian A, Gusarova N, Dobrenko N, et al. **Impact of adversarial examples on the efficiency of interpretation and use of information from high-tech medical images.** In: *24th Conference of Open Innovations Association (FRUCT)*, Moscow, Russian Federation. 2019:472–478 CrossRef
34. Hu L, Zhou DW, Guo XY, et al. **Adversarial training for prostate cancer classification using magnetic resonance imaging.** *Quant Imaging Med Surg* 2022;12:3276–87 CrossRef Medline
35. Han T, Nebelung S, Pedersoli F, et al. **Advancing diagnostic performance and clinical usability of neural networks via adversarial training and dual batch normalization.** *Nat Commun* 2021;12:4315. CrossRef Medline
36. Anastasiou T, Karagiorgou S, Petrou P, et al. **Towards robustifying image classifiers against the perils of adversarial attacks on artificial intelligence systems.** *Sensors (Basel)* 2022;22:6905 CrossRef Medline

# Deep Learning MRI Models for the Differential Diagnosis of Tumefactive Demyelination versus *IDH* Wild-Type Glioblastoma

 Gian Marco Conte,  Mana Moassefi, Paul A. Decker, Matthew L. Kosel, Christina B. McCarthy, Jessica A. Sagen, Yalda Nikanpour, Mahboubeh Fereidan-Esfahani, Michael W. Ruff, Fiorella S. Guido,  Heather K. Pump,  Terry C. Burns,  Robert B. Jenkins,  Bradley J. Erickson,  Daniel H. Lachance,  W. Oliver Tobin, and  Jeanette E. Eckel-Passow



## ABSTRACT

**BACKGROUND AND PURPOSE:** Diagnosis of tumefactive demyelination can be challenging. The diagnosis of indeterminate brain lesions on MRI often requires tissue confirmation via brain biopsy. Noninvasive methods for accurate diagnosis of tumor and nontumor etiologies allows for tailored therapy, optimal tumor control, and a reduced risk of iatrogenic morbidity and mortality. Tumefactive demyelination has imaging features that mimic *isocitrate dehydrogenase* wild-type glioblastoma (*IDHwt* GBM). We hypothesized that deep learning applied to postcontrast T1-weighted (T1C) and T2-weighted (T2) MRI can discriminate tumefactive demyelination from *IDHwt* GBM.

**MATERIALS AND METHODS:** Patients with tumefactive demyelination ( $n = 144$ ) and *IDHwt* GBM ( $n = 455$ ) were identified by clinical registries. A 3D DenseNet121 architecture was used to develop models to differentiate tumefactive demyelination and *IDHwt* GBM by using both T1C and T2 MRI, as well as only T1C and only T2 images. A 3-stage design was used: 1) model development and internal validation via 5-fold cross validation by using a sex-, age-, and MRI technology-matched set of tumefactive demyelination and *IDHwt* GBM, 2) validation of model specificity on independent *IDHwt* GBM, and 3) prospective validation on tumefactive demyelination and *IDHwt* GBM. Stratified area under the receiver operating curves (AUROCs) were used to evaluate model performance stratified by sex, age at diagnosis, MRI scanner strength, and MRI acquisition.

**RESULTS:** The deep learning model developed by using both T1C and T2 images had a prospective validation AUROC of 88% (95% CI: 0.82–0.95). In the prospective validation stage, a model score threshold of 0.28 resulted in 91% sensitivity of correctly classifying tumefactive demyelination and 80% specificity (correctly classifying *IDHwt* GBM). Stratified AUROCs demonstrated that model performance may be improved if thresholds were chosen stratified by age and MRI acquisition.

**CONCLUSIONS:** MRI can provide the basis for applying deep learning models to aid in the differential diagnosis of brain lesions. Further validation is needed to evaluate how well the model generalizes across institutions, patient populations, and technology, and to evaluate optimal thresholds for classification. Next steps also should incorporate additional tumor etiologies such as CNS lymphoma and brain metastases.

**ABBREVIATIONS:** AUROC = area under the receiver operator curve; CNSIDD = CNS inflammatory demyelinating disease; GBM = glioblastoma; *IDHwt* = *isocitrate dehydrogenase* wild-type; MOGAD = myelin oligodendrocyte glycoprotein antibody associated disorder; T1C = postcontrast T1-weighted; T2 = T2-weighted

The rate of diagnostic uncertainty, both before and after biopsy, for inflammatory tumefactive demyelinating lesions of the CNS is high.<sup>1-5</sup> It was shown that even on pathology

review, almost 30 % of 168 patients with biopsy-confirmed CNS inflammatory demyelinating disease (CNSIDD) were originally diagnosed incorrectly, leading to inappropriate therapy such as radiation and debulking in some patients.<sup>6</sup> Tumefactive MS is the most frequent cause of tumefactive demyelination and represents an extreme form of acute MS attack. Conventional imaging features of tumefactive MS mimic high-grade glioma, and particularly, glioblastoma, *isocitrate dehydrogenase* wild-type (*IDHwt* GBM).<sup>2,3</sup> These imaging features include size, mass effect, edema, contrast enhancement, diffusion restriction, and irregular lesion borders.

Most patients with tumefactive demyelination ultimately fulfill the 2017 McDonald criteria for the diagnosis of MS.<sup>7</sup> While tumefactive MS is the most frequent cause, other causes

Received October 28, 2024; accepted after revision December 30.

From the Departments of Radiology (G.M.C., M.M., Y.N., B.J.E.), Quantitative Health Sciences (P.A.D., M.L.K., J.E.E.-P.), Neurology (C.B.M., J.A.S., M.W.R., F.S.G., H.K.P., D.H.L., W.O.T.), Neurosurgery (T.C.B.), and Laboratory Medicine & Pathology (R.B.J.), and Center for Multiple Sclerosis and Autoimmune Neurology (W.O.T.), Mayo Clinic, Rochester, Minnesota; and Dell Medical School (M.F.-E.), University of Texas, Austin, Texas.

W. Oliver Tobin and Jeanette E. Eckel-Passow are co-senior authors.

This work was supported by the National Institutes of Health (R01NS113803).

Please address correspondence to Jeanette Eckel-Passow, 200 First St SW, Rochester, MN 55905; e-mail: Eckelpassow.jeanette@mayo.edu

 Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8645>

## SUMMARY

**PREVIOUS LITERATURE:** Diagnosis of tumefactive demyelination can be challenging and noninvasive methods for accurate diagnosis of tumor and nontumor etiologies is needed. The challenge of separating tumefactive demyelination from high-grade glioma has been addressed by using MRI radiomics, proton MR spectroscopy, nonstandard PET ligands, and CT. We hypothesized that deep learning applied to T1C and T2 MRI can discriminate tumefactive demyelination from *IDHwt* GBM.

**KEY FINDINGS:** We used a rigorous 3-stage experimental design, which included retrospective and prospective tumefactive demyelination and patients with *IDHwt* GBM. A deep learning model developed by using both T1C and T2 images had a prospective validation area under the receiver operator curve of 88% (95% CI: 0.82–0.95).

**KNOWLEDGE ADVANCEMENT:** We demonstrate that MRI can provide the basis for applying deep learning models to aid in the differential diagnosis of tumefactive demyelination and patients with *IDHwt* GBM. Further validation is needed to evaluate how the model generalizes across institutions, patient populations, and technology, and to evaluate optimal thresholds for classification.

of tumefactive demyelination include neuromyelitis optica spectrum disorder and myelin oligodendrocyte glycoprotein antibody associated disorder (MOGAD). While the incidence of tumefactive MS is rare (approximately 1–2 cases per 1000 cases of MS), the true incidence is likely larger given the known challenges with obtaining an accurate diagnosis.<sup>8</sup> Although MS is commonly thought of as a disease of the young, new onset inflammatory disease activity can occur at any age, with late onset inflammatory disease commonly being confused with tumor.<sup>9–11</sup> There are demographic differences in the presentation of *IDHwt* GBM and tumefactive MS; patients with new onset tumefactive MS tend to be younger, more likely to be women, less likely to present with seizure, and more likely to have other brain lesions or a history of MS. However, none of these characteristics are pathognomonic. The frequency of  $\kappa$  free light chains and oligoclonal bands in spinal fluid is lower in patients with tumefactive MS than typical MS, and the initial tumefactive lesion is the first demyelinating attack in more than two-thirds of patients.<sup>7,12</sup> Notwithstanding its severity, tumefactive MS is quite treatable. Even in patients with very severe disease, treatment with cyclophosphamide has been shown to be effective.<sup>13</sup> Conversely, standard of care for *IDHwt* GBM is maximal safe resection followed by chemoradiation.<sup>14</sup> Thus, misdiagnosis of tumefactive MS as *IDHwt* GBM can lead to radiation therapy and surgery, which worsens outcome for patients with tumefactive MS.<sup>15</sup>

Even though surgery for patients with tumefactive MS leads to worse outcome, the diagnosis of indeterminate MRI brain lesions often requires tissue confirmation via brain biopsy. Noninvasive methods for accurate diagnosis of tumor and nontumor etiologies for brain lesions will allow for tailored therapy, optimal tumor control, and a reduced risk of iatrogenic morbidity and mortality, such as from surgery or irradiation in tumefactive MS.<sup>6</sup> MRI features, including ADC imaging features and T2 hypointense rims can help differentiate tumefactive demyelination from *IDHwt* GBM.<sup>16–20</sup> While these characteristics can help with differential diagnosis, there is overlap in MRI appearances (Fig 1). To overcome this issue, deep learning MRI analyses have been performed.<sup>21–24</sup> Deep learning is a subset of artificial intelligence and machine learning and does not require a priori identification of

features to use as candidate predictor variables. Instead, deep learning models automatically select features from an input (such as an MR image) to consider as predictors in the model.

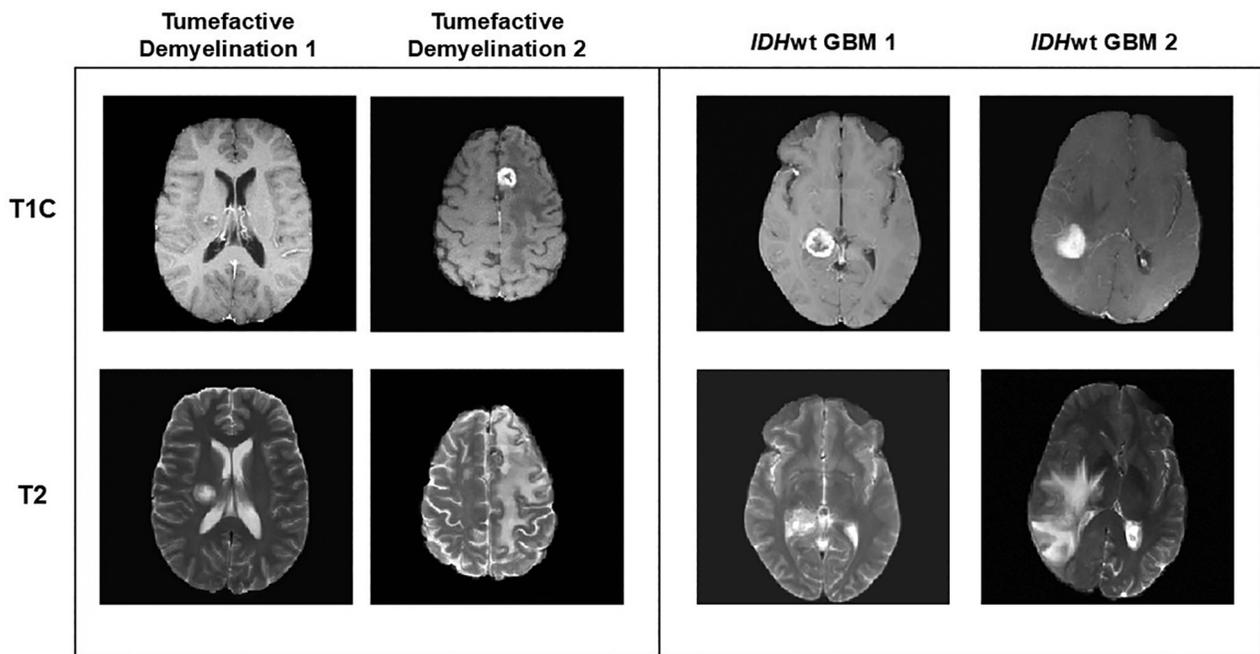
We hypothesized that deep learning applied to postcontrast T1-weighted (T1C), and T2-weighted (T2) images could help with the differential diagnosis of tumefactive demyelination versus *IDHwt* GBM. We demonstrate the importance of careful experimental design and prospective validation, as well as the usefulness of evaluating stratified versions of the area under the receiver operator curve (AUROC) to understand model performance.

## MATERIALS AND METHODS

### Patient Selection

The institutional review board at the local institution approved this study. Patients consented for use of their biospecimens or data. The study was performed by using a 3-stage experimental design based on diagnosis date. As a referral center and large health system, all 3 stages included MRIs that were obtained from many hospitals.

**Stage 1: Model Development and Internal Validation.** A total of 110 patients with tumefactive demyelination diagnosed before January 1, 2020, were identified by the Center for Multiple Sclerosis and Autoimmune Neurology. The diagnosis of CNSIDD was confirmed by review of at least 1 fellowship-trained multiple sclerosis neurologist. Fourteen (13%) of the patients with tumefactive demyelination underwent a biopsy. All cases were tumefactive MS except for 4 cases of MOGAD. The term tumefactive demyelination is used to refer to all patients with CNSIDD in this article. A total of 386 patients with *IDHwt* GBM were identified from our clinical neuro-oncology registry and had a confirmed pathologic diagnosis before January 1, 2020. *IDH* status was assessed for clinical or research purposes by using immunohistochemistry or sequencing. Inclusion criteria were age  $\geq 18$  years and availability of preoperative T1C and T2 MR images. CNSIDD cases were also required to have a lesion with a minimum transverse diameter  $\geq 10$  mm and fewer than 5 tumefactive lesions. Exclusion criteria were previous history of brain surgery or radiation therapy for the treatment of a brain tumor and low-quality MRI due to severe motion artifacts as determined by manual visual inspection. To develop an MRI-based



**FIG 1.** T1C and T2 images for 2 patients with tumefactive demyelination and 2 patients with *IDHwt* GBM.

deep learning model in stage 1, 110 patients with *IDHwt* GBM were matched to 110 patients with tumefactive demyelination. We matched exactly on sex and MRI acquisition and subsequently performed optimal matching on age at diagnosis and index MRI year.

**Stage 2: Validation on an Independent Series of Retrospective *IDHwt* GBM to Evaluate Model Specificity.** The remaining 276 patients with *IDHwt* GBM who were diagnosed before January 1, 2020, and who were not used in stage 1 were used for validation in stage 2.

**Stage 3: Validation on a Prospective Series of Tumefactive Demyelination and *IDHwt* GBM to Evaluate Sensitivity and Specificity.** Thirty-four patients with tumefactive demyelination and 69 patients with *IDHwt* GBM diagnosed or consented on or after January 1, 2020, were used in stage 3. There was 1 patient with tumefactive demyelination consented after January 1, 2020; however, their diagnosis was in 2009. Patients with tumefactive demyelination included 23 relapsing-remitting MS, 10 focal cerebral demyelinating, and 1 MOGAD. Three (9%) of the patients with tumefactive demyelination underwent a biopsy.

### **MRI Preprocessing**

Each MRI examination was processed independently. HD-GLIO-AUTO<sup>25,26</sup> was used to perform image registration and brain extraction. The preprocessing pipeline implemented in the Federated Tumor Segmentation platform<sup>27</sup> was applied to register each MRI volume to a common anatomic space, by using the SRI24 atlas<sup>28</sup> and an isotropic voxel resolution of 1 mm<sup>3</sup>. This ensured a consistent data shape of ( $x = 240, y = 240, z = 155$ ), where  $x$  and  $y$  denote the axial axes and  $z$  the number of slices. Each patient's MRI signal intensities were standardized by subtracting the mean and dividing by the standard deviation.

### **Deep Learning Architecture**

To differentiate tumefactive demyelination from *IDHwt* GBM, a 3D DenseNet121 architecture was implemented in the MONAI framework (v. 1.1.0)<sup>29</sup> by using default settings. Three different models were developed by considering different inputs: 1) 2-channel input composed of the T1C and T2 3D volumes, 2) single-channel model composed of the T1C 3D volumes, and 3) a single-channel model composed of the T2 3D volumes. The DenseNet121 architecture was trained by using a priori chosen hyperparameters: 650 epochs, batch size of 16, cosine annealing learning rate scheduler starting at  $10^{-3}$ , binary cross-entropy loss, and AdamW optimizer. Data augmentation included random flipping, translation, rotation, and scaling, with 50% probability. The stopping rule was chosen as the first epoch across 3 sequential epochs with a difference in cross-entropy loss  $< 0.02$ . Python (v. 3.9.15), PyTorch (v.1.13.1), and MONAI (v. 1.1.0) libraries were used.

### **Analysis and Model Performance**

The experimental design is summarized in the Supplemental Data and consisted of a model development stage that utilized internal validation and 2 independent validation stages. To develop a final model, the deep learning architecture was applied to the full data set. Three sets of 5-fold cross validation were used to obtain an unbiased estimate of prediction performance.<sup>30</sup> For each cross-validation fold, a model was developed by using 80% of the data and the AUROC for the corresponding model was calculated on the 20% left-out test set. The average AUROC across the 15 left-out test sets, and the corresponding 95% CI, were calculated. In the prospective validation stage (stage 3), AUROC and 95% CIs were calculated by using the DeLong method. Stratified AUROCs were calculated in the prospective validation stage to evaluate if model performance was associated with demographic or MRI characteristics. A stratified AUROC is the AUROC estimated separately for each

**Clinical and MRI characteristics of patients used in stage 1, stage 2, and stage 3**

	Stage 1		Stage 2	Stage 3	
	Tumefactive Demyelination	IDHwt GBM	IDHwt GBM	Tumefactive Demyelination	IDHwt GBM
	(n = 110)	(n = 110)	(n = 276)	(n = 34)	(n = 69)
Index MRI year					
Median (min, max)	2012 (1998, 2019)	2015 (2003, 2019)	2015 (2001, 2019)	2021 (2009, <sup>a</sup> 2023)	2020 (2020, 2021)
Sex					
Man	44 (40.0%)	44 (40.0%)	198 (71.7%)	13 (38.2%)	42 (60.9%)
Age at diagnosis					
Median (min, max)	40 (19, 70)	51 (19, 71)	62 (22, 93)	34.2 (18, 72)	65.3 (19, 83)
MRI Manufacturer					
GE Healthcare	55 (50.0%)	54 (49.1%)	141 (51.1%)	8 (23.5%)	26 (37.6%)
Philips Healthcare	10 (9.1%)	4 (3.6%)	16 (5.8%)	1 (2.9%)	3 (4.4%)
Siemens	41 (37.3%)	47 (42.7%)	101 (36.6%)	22 (64.7%)	38 (55.1%)
Other	4 (3.6%)	5 (4.5%)	18 (6.5%)	3 (8.8%)	2 (2.9%)
MRI field strength					
1.5	91 (82.7%)	85 (77.3%)	165 (59.8%)	15 (44.1%)	34 (49.3%)
3	17 (15.5%)	23 (20.9%)	100 (36.2%)	17 (50.0%)	34 (49.3%)
Other	0 (0.0%)	0 (0.0%)	1 (0.4%)	2 (5.9%)	0 (0.0%)
Missing	2 (1.8%)	2 (1.8%)	10 (3.6%)	0 (0.0%)	1 (1.4%)
T1C MRI acquisition					
2D	83 (75.4%)	83 (75.5%)	96 (34.8%)	16 (47.1%)	25 (36.8%)
3D	20 (18.2%)	27 (24.5%)	166 (60.1%)	18 (52.9%)	43 (63.2%)
Missing	7 (6.4%)	0 (0.0%)	14 (5.1%)	0 (0.0%)	1 (1.4%)
T2 MRI acquisition					
2D	102 (92.7%)	108 (98.2%)	257 (93.1%)	33 (97.1%)	57 (83.8%)
3D	0 (0.0%)	2 (1.8%)	6 (2.2%)	1 (2.9%)	11 (16.2%)
Missing	8 (7.3%)	0 (0.0%)	13 (4.7%)	0 (0.0%)	1 (1.4%)

**Note:**—Stage 1 and stage 2 included patients diagnosed before January 1, 2020, while stage 3 included patients diagnosed on or after January 1, 2020.

<sup>a</sup> There was 1 patient with tumefactive MS who was consented after January 1, 2020; however, their diagnosis was in 2009. All other patients were diagnosed after January 1, 2020.

stratum, eg, when stratifying by sex the AUROC was calculated among women and men separately. Occlusion maps were generated to identify which regions of the brain contributed model predictions, as described in the Supplemental Data.

## RESULTS

### Stage 1: Model Development and Internal Validation

A total of 110 patients with tumefactive demyelination met the study inclusion and exclusion criteria and 110 patients with IDHwt GBM were matched to tumefactive demyelination on sex, age at diagnosis, index MRI date, and 2D/3D MRI acquisition (Table). Due to matching to tumefactive demyelination, 40% of the patients were men in both groups. The median age at diagnosis for tumefactive demyelination and IDHwt GBM was 40 and 51, respectively. While we aimed to match on age at diagnosis to prevent the deep learning model from predicting age, it was challenging to find young patients with IDHwt GBM. To identify potential experimental design bias, MRI technology was compared across groups before performing deep learning analyses. GE Healthcare and Siemens were the most common manufacturers, 1.5T was the most common field strength, and 2D was the most common acquisition for both T1C and T2 images (Table).

Three models were developed by using different sets of MRI inputs: 1) both T1C and T2, 2) only T1C, and 3) only T2 images. The average AUROC across the 15 cross-validation folds was 0.86 (95% CI: 0.83–0.89), 0.83 (95% CI: 0.78–0.87), and 0.83 (95% CI: 0.80–0.87), for models that used both T1C and T2, T1C only, and T2 only, respectively.

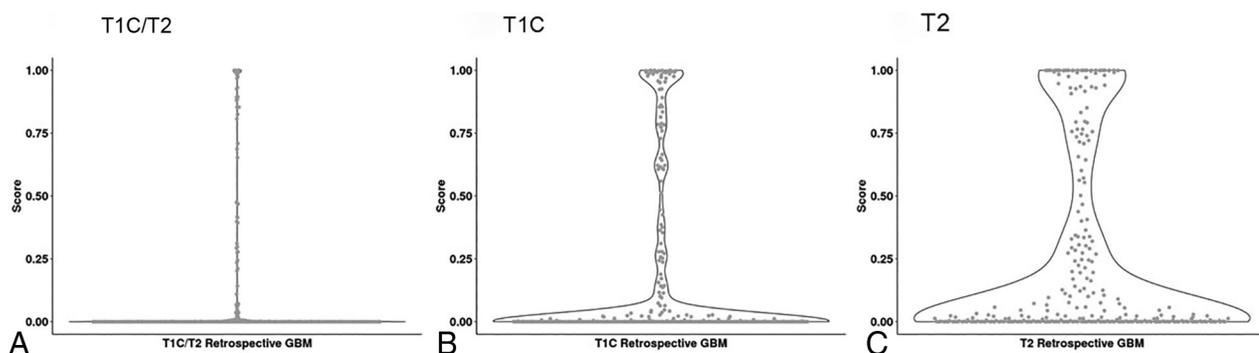
### Stage 2: Validation on an Independent Series of Retrospective Patients with IDHwt GBM

An independent set of 276 patients with IDHwt GBM diagnosed before January 1, 2020, were utilized to evaluate the specificity of the models developed in stage 1 (Table). The patients with IDHwt GBM were 72% men, reflecting the known higher prevalence of IDHwt GBM in men, and the median age was 62. MRIs were primarily processed on GE Healthcare and Siemens and with 1.5T field strength. T1 acquisition was 63% 3D in stage 2, versus 25% 3D in stage 1.

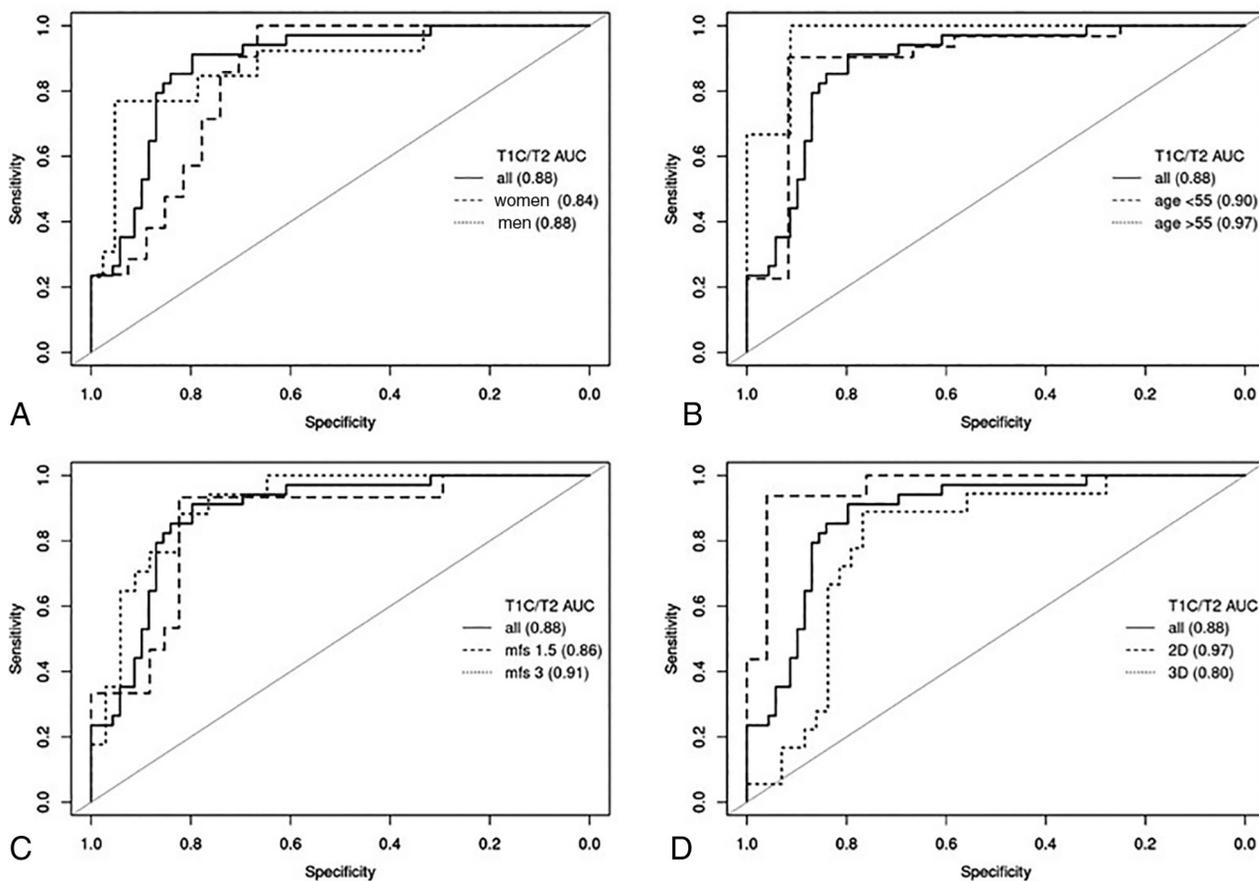
Fig 2 displays the distribution of scores from the 276 patients with IDHwt GBM from the deep learning models derived by using both T1C and T2, only T1C, and only T2 images. A score near 1 denotes higher confidence of being tumefactive demyelination, whereas a score near zero denotes higher confidence of being IDHwt GBM. Of the 276 patients with IDHwt GBM, 86%, 77%, and 69% had a score <0.25 from the models derived by using both T1C and T2, T1C only, and T2 only, respectively.

### Stage 3: Validation on a Prospective Series of Patients with Tumefactive Demyelination and IDHwt GBM

An independent set of 34 patients with tumefactive demyelination and 69 patients with IDHwt GBM diagnosed on or after January 1, 2020, were utilized to validate prospectively the models developed in stage 1 (Table). Of the 34 patients with tumefactive demyelination, 33 were diagnosed between 2020 and 2023 (1 patient was diagnosed in 2009), whereas the 69 patients with IDHwt GBM were diagnosed between 2020 and 2021. The sex and age distributions reflect the known characteristics of tumefactive demyelination and



**FIG 2.** Distribution of scores from 276 patients with *IDHwt* GBM for the models derived from using (A) both T1C and T2, (B) only T1C, and (C) only T2 MRI. The models were developed to predict tumefactive demyelination; thus, patients with *IDHwt* GBM are expected to have scores near zero and patients with tumefactive demyelination are expected to have scores near 1.

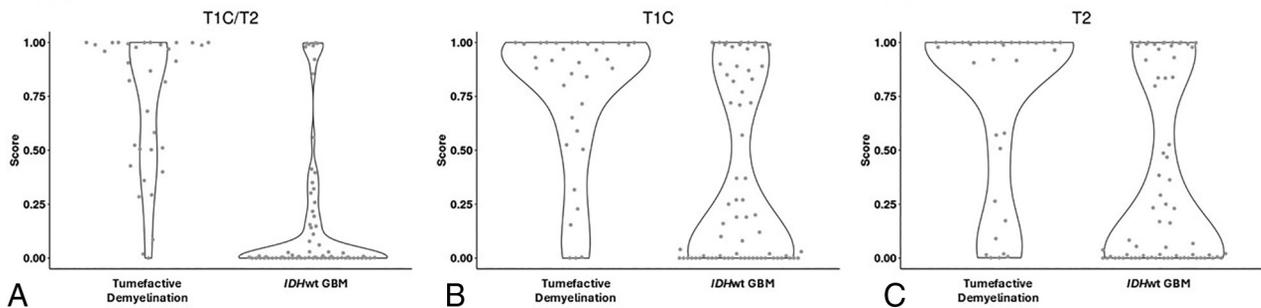


**FIG 3.** Stratified AUROCs from the prospective validation for the model derived by using both T1C and T2 images. The AUROCs are stratified by (A) sex, (B) age at diagnosis, (C) MRI field strength, and (D) MRI acquisition.

*IDHwt* GBM. MRI characteristics changed over time; however, these changes reflect how technology changes over time and reiterates the importance of prospective validation. While before 2020 GE Healthcare was the most common manufacturer, after 2020, Siemens was more common. Likewise, field strength and T1C acquisition changed over time.

The models developed by using both T1C and T2, only T1C, and only T2 images had validation AUROCs of 0.88 (95% CI: 0.82–0.95), 0.75 (95% CI: 0.66–0.84), and 0.77 (95% CI: 0.67–0.86), respectively. Stratified receiver operator curves were created, and

the corresponding AUROCs were estimated to evaluate if model performance was associated with demographic (age and sex) or MRI (acquisition and field strength) characteristics. This is particularly important because demographic and MRI characteristics can change over time and thus were different between the patients that the models were developed on versus the patients in the prospective validation stage (Table). Fig 3 provides stratified AUROCs by sex, age at diagnosis, MRI scanner strength, and MRI acquisition parameters for the model developed by using both T1C and T2 images. Due to the small sample size of the stratified analyses in



**FIG 4.** Distribution of scores from the prospective validation that consisted of 34 patients with tumefactive demyelination and 69 patients with *IDHwt* GBM, for the models derived by using (A) both T1C and T2, (B) only T1C, and (C) only T2 MRI.

the prospective validation (stage 3), one must be careful to not over interpret the stratified AUROCs; thus, the corresponding results should be considered descriptive. Overall, most of the stratified AUROCs fall within the 95% CI of the AUROC calculated from all the patients in the prospective validation (AUROC = 0.88, 95% CI: 0.82–0.95). Figure 3B and -D show that there might be a difference in model performance by age (<55 versus >55 years old) and MRI acquisition (2D versus 3D) if stratified-specific thresholds are used to classify patients; however, a larger validation set is necessary to confirm these observations and to be adequately powered to perform formal statistical comparisons. Specifically, Fig 3B demonstrates where an age-specific threshold for classifying tumefactive demyelination versus *IDHwt* GBM could possibly improve performance across all patients, ie, the AUROC for both age groups is larger than the overall AUROC. Figure 3D demonstrates that model performance for 2D MRIs can be improved with an acquisition-specific threshold; however, this comes at a cost to the performance of 3D MRIs. It is not surprising that we observed that 2D had improved performance because 2D was used in 75% of the training set (stage 1). Supplemental Data contain stratified AUROCs for the models developed by using only T1C and only T2 images, respectively, and demonstrate similar observations.

Figure 4 shows the distribution of scores for the prospective validation subjects derived from each of the 3 deep learning models. To perform classification, a threshold needs to be chosen such that a score larger than the threshold is classified as tumefactive demyelination and a score less than the threshold is classified as *IDHwt* GBM. In determining a threshold there is an inherent trade-off between sensitivity (ie, correctly predicting a tumefactive demyelination patient to have tumefactive demyelination) and specificity (ie, correctly predicting a *IDHwt* GBM patient to have *IDHwt* GBM). The Supplemental Data provide sensitivity and specificity estimates associated with all possible thresholds in the prospective validation set. As an example, if the goal is to develop a classification model with at least 90% sensitivity for tumefactive demyelination, then a threshold of 0.28 would correspond to a sensitivity of 91% and specificity of 80% in the prospective validation.

Figure 4A demonstrates that there are patients with tumefactive demyelination who have scores near zero and thus will be misclassified as *IDHwt* GBM. Likewise, there are patients with *IDHwt* GBM who have scores near 1 and thus will be misclassified

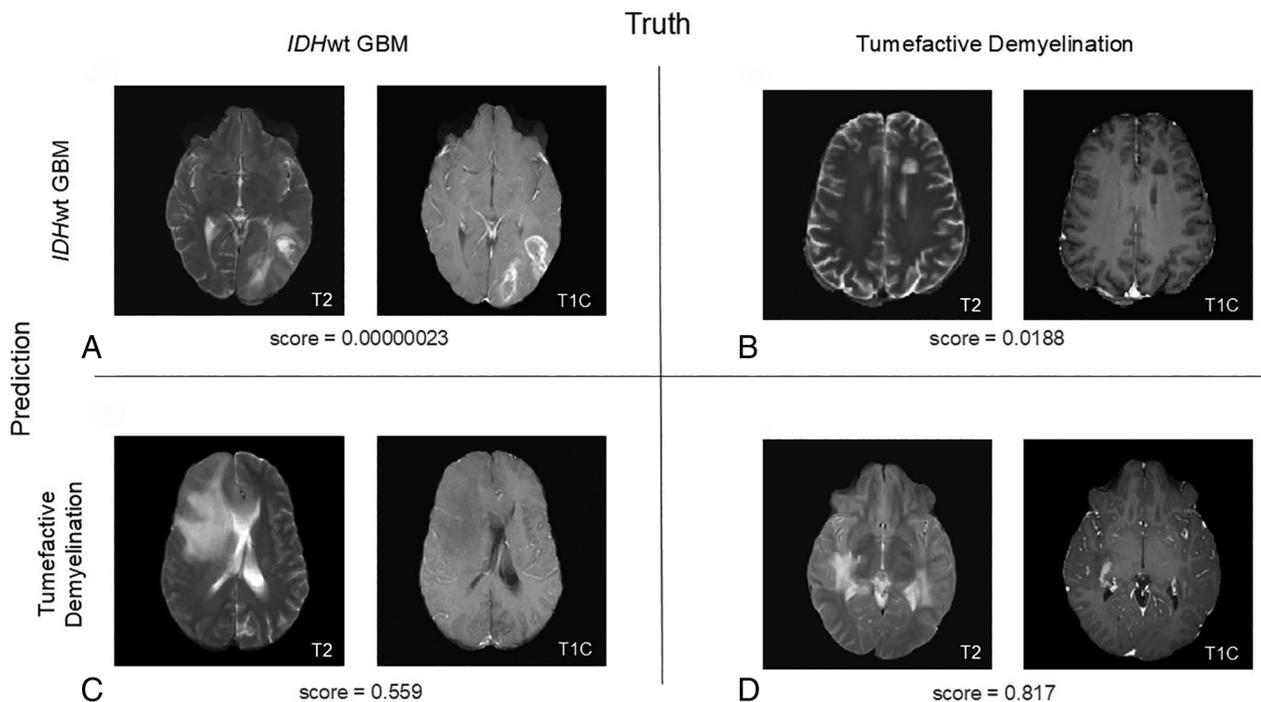
as tumefactive demyelination. Figure 5 shows example T1C and T2 images for both correctly classified patients and for patients who would likely be misclassified. The Supplemental Data provide occlusion maps from the same 2D slices that are shown in Fig 5. The occlusion maps, because they only show a 2D slice, represent a very small piece of the overall data that was used in the 3D models and thus, might not be representative of the final model prediction. The Supplemental Data provide a different 2D slice from the same set of patients.

## DISCUSSION

The challenge of separating tumefactive demyelination from high-grade glioma has been addressed by using MRI radiomics,<sup>31,32</sup> proton MR spectroscopy,<sup>33</sup> nonstandard PET ligands,<sup>34–36</sup> and CT.<sup>37</sup> Techniques to distinguish these entities should be available at the time of the initial MRI. Once a mass suspicious of *IDHwt* GBM is identified, there is often a push to proceed quickly to remove it. While an *IDHwt* GBM is much more common than tumefactive demyelination, treating physicians must consciously consider this rarer alternative because disability and clinical outcome for patients with tumefactive demyelination is related to the amount of tissue that is removed when surgery is performed.

Herein, we implemented a 3-stage study design that included 144 patients with tumefactive MS and 455 patients with *IDHwt* GBM and utilized deep learning methods. Using both T2 and T1C images, the deep learning model had a prospective validation AUROC of 88%. To determine a threshold for classification, we demonstrated an example where we optimized sensitivity of correctly predicting tumefactive demyelination. The rationale was that performing a total or subtotal resection on a tumefactive demyelination lesion would lead to substantial deficit, without a treatment benefit. Conversely, initially misdiagnosing an *IDHwt* GBM would typically allow for a systemic treatment such as corticosteroids or plasma exchange.<sup>6</sup> For this example, we chose a threshold that obtained 90% sensitivity of correctly classifying tumefactive demyelination, which resulted in 80% specificity.

While the diagnostic agreement between a radiologist and the MRI-based models can be compared, such analyses will result in a biased estimate if the groups do not reflect the actual prevalence or characteristics of each disease. A more relevant and accurate metric is the biopsy rate among patients with tumefactive demyelination, which provides an estimate of real-life clinical diagnostic



**FIG 5.** Example TIC and T2 MRI for (A) a patient with *IDHwt* GBM who was classified correctly, (B) a patient with tumefactive demyelination who was classified incorrectly, (C) a patient with *IDHwt* GBM who was classified incorrectly, and (D) a patient with tumefactive demyelination who was classified correctly. The score was derived from the model that used both TIC and T2 images.

uncertainty. In our analysis cohorts, 13% of the retrospective and 9% of the prospective patients with tumefactive demyelination underwent biopsy. Other published cohorts of patients with tumefactive demyelination report similar or higher rates of biopsy; thus, confirming diagnostic uncertainty in this cohort across institutions.<sup>1,2,5</sup>

Experimental design is critically important so that clinical and technical artifacts do not bias the prediction models.<sup>38</sup> This study utilized a rigorous 3-stage experimental design where each stage included MRIs that were generated at multiple institutions. In stage 1, matching was utilized to reduce the chances of developing a biased model, ie, a model that is confounded with clinical features (eg, age and sex) or technical artifacts such as MRI technology.<sup>38</sup> Specifically, we did not want to develop a model that was a surrogate of age, sex, or MRI technology. In stage 2, validation was performed on a large independent set of retrospective patients with *IDHwt* GBM. Stage 3 entailed prospective validation. Due to changes in MRI technology and acquisition protocols over time, as well as changes in the tumefactive demyelination clinical diagnostic criteria, it is imperative to evaluate how a model performs in prospective patients. We acknowledge that the prospective validation had a limited sample size. Tumefactive demyelination is a rare disease, and the sample size analyzed herein is one of the largest sets published to date. The overall design and validation approach was further strengthened by utilizing stratified AUROCs to evaluate the impact of age, sex, and MRI technology on model performance. In future validation efforts, the impact of additional features on model performance, such as in-plane spatial resolution, should similarly be evaluated.

The analysis approach utilized a DenseNet121 architecture. This architecture was chosen due to its efficient use of parameters

and the ability to reuse features, which results in a compact model.<sup>39</sup> There is virtually an infinite number of deep learning architectures available, and a full comparison of approaches is beyond the scope of this study. Additionally, T1C and T2 MRI sequences were used as model input, because of consistency of data availability. Features that can be identified on T1C and T2 images, such as Baló concentric sclerosis and open ring sign, were not readily available for all patients and thus were not manually included as predictors in the model. However, the 3D deep learning model would be capable of identifying and using these features if they were determined to be important for prediction. ADC imaging sequences were not included in the model due to the variability of ADC source images and significant changes that occurred in ADC acquisition parameters over time. Imaging features that were added to the 2024 iteration of the MS diagnostic criteria, including paramagnetic rim sign and central vein sign, were not available for many of the patients due to the retrospective nature of the design.

## CONCLUSIONS

Overall, this study demonstrates that T1C and T2 MRI can provide the basis for applying deep learning models to aid in the differential diagnosis of brain lesions. We do not propose that an MRI model will be used in isolation. Instead, we suggest that it could be used as one of many tests to help with differential diagnosis, which would ideally help reduce the number of biopsies done on patients with tumefactive demyelination because disability and clinical outcome is related to the amount of tissue that is removed when surgery is performed. Deep learning models may be most impactful in practice settings where subspecialty clinical

expertise may not be available, which is particularly important for a rare disease such as tumefactive demyelination. Even at a referral hospital with expertise in this area, 9% of the patients analyzed herein who were diagnosed with tumefactive demyelination since 2020 underwent a biopsy. While the models presented herein demonstrate feasibility, further validation is needed to choose and validate a model score threshold, evaluate generalizability of the model, and to evaluate how clinical, demographic, and MRI features affect model performance. An important challenge to clinical adoption of deep learning models is their robustness to changes in institution, patient population, and MRI technology. Future work should also include brain metastases and CNS lymphoma; both can have similar appearances on MRI to tumefactive demyelination and *IDHwt* GBM.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

- Balloy G, Pelletier J, Suchet L, et al; Société Francophone de la Sclérose en Plaques. **Inaugural tumor-like multiple sclerosis: clinical presentation and medium-term outcome in 87 patients.** *J Neurol* 2018;265:2251–59 CrossRef Medline
- Frankl S, Viaeane A, Vossough A, et al. **Solitary tumefactive demyelinating lesions in children: clinical and magnetic resonance imaging features, pathologic characteristics, and outcomes.** *Pediatr Neurol* 2024;157:141–50 CrossRef Medline
- Pervin I, Ramanathan S, Cappelen-Smith C, et al. **Clinical and radiological characteristics and outcomes of patients with recurrent or relapsing tumefactive demyelination.** *Mult Scler Relat Disord* 2024;82:105408 CrossRef Medline
- Tobin WO, Meyer FB, Keegan BM. **Diagnostic yield and safety of cerebellar and brainstem parenchymal biopsy.** *World Neurosurg* 2015;84:1973–76 CrossRef Medline
- Tremblay MA, Villanueva-Meyer JE, Cha S, et al. **Clinical and imaging correlation in patients with pathologically confirmed tumefactive demyelinating lesions.** *J Neurol Sci* 2017;381:83–87 CrossRef Medline
- Lucchinetti CF, Gavrilova RH, Metz I, et al. **Clinical and radiographic spectrum of pathologically confirmed tumefactive multiple sclerosis.** *Brain* 2008;131:1759–75 CrossRef Medline
- Fereidan-Esfahani M, Decker PA, Weigand SD, et al. **Defining the natural history of tumefactive demyelination: a retrospective cohort of 257 patients.** *Ann Clin Transl Neurol* 2023;10:1544–55 CrossRef Medline
- Fereidan-Esfahani M, Decker PA, Eckel Passow JE, et al. **Population-based incidence and clinico-radiological characteristics of tumefactive demyelination in Olmsted County, Minnesota, United States.** *Eur J Neurol* 2022;29:782–89 CrossRef Medline
- Mouresan EF, Mentessidou E, Berglund A, et al. **Clinical characteristics and long-term outcomes of late-onset multiple sclerosis: a Swedish nationwide study.** *Neurology* 2024;102:e208051 CrossRef Medline
- Tobin WO, Costanzi C, Guo Y, et al. **Clinical-radiological-pathological spectrum of central nervous system-idiopathic inflammatory demyelinating disease in the elderly.** *Mult Scler* 2017;23:1204–13 CrossRef Medline
- Lin X, Yu WY, Liauw L, et al. **Clinicoradiologic features distinguish tumefactive multiple sclerosis from CNS neoplasms.** *Neurol Clin Pract* 2017;7:53–64 CrossRef Medline
- Banks SA, Willrich MAV, Eckel-Passow JE, et al. **Cerebrospinal fluid kappa free light chains in patients with tumefactive demyelination.** *Clin Chem Lab Med* 2024;62:e189–92 CrossRef Medline
- Fereidan-Esfahani M, Tobin WO. **Cyclophosphamide in treatment of tumefactive multiple sclerosis.** *Mult Scler Relat Disord* 2021;47:102627 CrossRef Medline
- Horbinski C, Nabors LB, Portnow J, et al. **NCCN Guidelines(R) Insights: Central Nervous System Cancers, Version 2.2022.** *J Natl Compr Canc Netw* 2023;21:12–20 CrossRef Medline
- Miller RC, Lachance DH, Lucchinetti CF, et al. **Multiple sclerosis, brain radiotherapy, and risk of neurotoxicity: the Mayo Clinic experience.** *Int J Radiat Oncol Biol Phys* 2006;66:1178–86 CrossRef Medline
- Abou Zeid N, Pirko I, Erickson B, et al. **Diffusion-weighted imaging characteristics of biopsy-proven demyelinating brain lesions.** *Neurology* 2012;78:1655–62 CrossRef Medline
- Mabray MC, Cohen BA, Villanueva-Meyer JE, et al. **Performance of apparent diffusion coefficient values and conventional MRI features in differentiating tumefactive demyelinating lesions from primary brain neoplasms.** *AJR Am J Roentgenol* 2015;205:1075–85 CrossRef Medline
- Cacciaguerra L, Morris P, Tobin WO, et al. **Tumefactive demyelination in MOG Ab-associated disease, multiple sclerosis, and AQP-4-IgG-positive neuromyelitis optica spectrum disorder.** *Neurology* 2023;100:e1418–e1432 CrossRef Medline
- Koelblinger C, Fruehwald-Pallamar J, Kubin K, et al. **Atypical idiopathic inflammatory demyelinating lesions (IIDL): conventional and diffusion-weighted MR imaging (DWI) findings in 42 cases.** *Eur J Radiology* 2013;82:1996–2004 CrossRef Medline
- Schwartz KM, Erickson BJ, Lucchinetti C. **Pattern of T2 hypointensity associated with ring-enhancing brain lesions can help to differentiate pathology.** *Neuroradiology* 2006;48:143–49 CrossRef Medline
- Zhang Y, Liang K, He J, et al. **Deep learning with data enhancement for the differentiation of solitary and multiple cerebral glioblastoma, lymphoma, and tumefactive demyelinating lesion.** *Front Oncol* 2021;11:665891 CrossRef Medline
- Ryu WS, Kang YR, Noh YG, et al. **Acute infarct segmentation on diffusion-weighted imaging using deep learning algorithm and RAPID MRI.** *J Stroke* 2023;25:425–29 CrossRef Medline
- Lee J, Burkett BJ, Min HK, et al. **Synthesizing images of tau pathology from cross-modal neuroimaging using deep learning.** *Brain* 2024;147:980–95 CrossRef Medline
- Wang W, Zhao Y, Teng L, et al. **Neuropathologist-level integrated classification of adult-type diffuse gliomas using deep learning from whole-slide pathological images.** *Nat Commun* 2023;14:6359 CrossRef Medline
- Isensee F, Jaeger PF, Kohl SAA, et al. **nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation.** *Nat Methods* 2021;18:203–11 CrossRef Medline
- Isensee F, Schell M, Pfleger I, et al. **Automated brain extraction of multisequence MRI using artificial neural networks.** *Hum Brain Mapp* 2019;40:4952–64 CrossRef Medline
- Pati S, Baid U, Edwards B, et al. **The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research.** *Phys Med Biol* 2022;67: CrossRef Medline
- Rohlfing T, Zahr NM, Sullivan EV, et al. **The SRI24 multichannel atlas of normal adult human brain structure.** *Hum Brain Mapp* 2010;31:798–819 CrossRef Medline
- Cardoso MJ, Li W, Brown R, et al. **MONAI: An Open-Source Framework for Deep Learning in Healthcare.** *arXiv.org* 2022. <https://doi.org/10.48550/arXiv.2211.02701>. Accessed May 1, 2024
- Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* New York: Springer-Verlag; 2009.
- Kolakshyapati M, Hashizume A, Ochi K, et al. **Usefulness of histogram-profile analysis in ring-enhancing intracranial lesions.** *World Neurosurg* 2019;131:e226–36 CrossRef Medline
- Verma RK, Wiest R, Locher C, et al. **Differentiating enhancing multiple sclerosis lesions, glioblastoma, and lymphoma with dynamic texture parameters analysis (DTPA): a feasibility study.** *Med Phys* 2017;44:4000–08 CrossRef Medline

33. Ikeguchi R, Shimizu Y, Abe K, et al. **Proton magnetic resonance spectroscopy differentiates tumefactive demyelinating lesions from gliomas.** *Mult Scler Relat Disord* 2018;26:77–84 CrossRef Medline
34. Hashimoto S, Inaji M, Nariai T, et al. **Usefulness of [(11)C] methionine PET in the differentiation of tumefactive multiple sclerosis from high grade astrocytoma.** *Neurol Med Chir (Tokyo)* 2019; 59:176–83 CrossRef Medline
35. Kebir S, Rauschenbach L, Weber M, et al. **Machine learning-based differentiation between multiple sclerosis and glioma WHO II degrees -IV degrees using O-(2-[18F] fluoroethyl)-L-tyrosine positron emission tomography.** *J Neurooncol* 2021;152:325–32 CrossRef Medline
36. Yasuda S, Yano H, Kimura A, et al. **Frontal tumefactive demyelinating lesion mimicking glioblastoma differentiated by methionine positron emission tomography.** *World Neurosurg* 2018;119:244–48 CrossRef Medline
37. Kim DS, Na DG, Kim KH, et al. **Distinguishing tumefactive demyelinating lesions from glioma or central nervous system lymphoma: added value of unenhanced CT compared with conventional contrast-enhanced MR imaging.** *Radiology* 2009;251:467–75 CrossRef Medline
38. Ransohoff DF. **Bias as a threat to the validity of cancer molecular-marker research.** *Nat Rev Cancer* 2005;5:142–49 CrossRef Medline
39. Huang G, Liu Z, van der Maaten L, et al. **Densely connected convolutional networks.** *arXiv.org* 2016. <https://doi.org/10.48550/arXiv.1608.06993>. Accessed May 1, 2024

# Image-Based Search in Radiology: Identification of Brain Tumor Subtypes within Databases Using MRI-Based Radiomic Features

 Marc von Reppert,  Saahil Chadha, Klara Willms,  Arman Avesta,  Nazanin Maleki,  Tal Zeevi, Jan Lost,  Niklas Tillmanns, Leon Jekel, Sara Merkaj,  MingDe Lin,  Karl-Titus Hoffmann,  Sanjay Aneja, and Mariam S. Aboian, for the IBSR Consortium



## ABSTRACT

**BACKGROUND AND PURPOSE:** Existing neuroradiology reference materials do not cover the full range of primary brain tumor presentations, and text-based medical image search engines are limited by the lack of consistent structure in radiology reports. To address this, an image-based search approach is introduced here, leveraging an institutional database to find reference MRIs visually similar to presented query cases.

**MATERIALS AND METHODS:** Two hundred ninety-five patients (mean age and standard deviation,  $51 \pm 20$  years) with primary brain tumors who underwent surgical and/or radiotherapeutic treatment between 2000 and 2021 were included in this retrospective study. Semiautomated convolutional neural network–based tumor segmentation was performed, and radiomic features were extracted. The data set was split into reference and query subsets, and dimensionality reduction was applied to cluster reference cases. Radiomic features extracted from each query case were projected onto the clustered reference cases, and nearest neighbors were retrieved. Retrieval performance was evaluated by using mean average precision at  $k$ , and the best-performing dimensionality reduction technique was identified. Expert readers independently rated visual similarity by using a 5-point Likert scale.

**RESULTS:** t-Distributed stochastic neighbor embedding with 6 components was the highest-performing dimensionality reduction technique, with mean average precision at 5 ranging from 78%–100% by tumor type. The top 5 retrieved reference cases showed high visual similarity Likert scores with corresponding query cases (76% ‘similar’ or ‘very similar’).

**CONCLUSIONS:** We introduce an image-based search method for exploring historical MR images of primary brain tumors and fetching reference cases closely resembling queried ones. Assessment involving comparison of tumor types and visual similarity Likert scoring by expert neuroradiologists validates the effectiveness of this method.

**ABBREVIATIONS:** A/O = astrocytoma and oligodendroglioma WHO CNS grades 2–3; CNN = convolutional neural network; G/A = glioblastoma and astrocytoma WHO CNS grade 4; ICC = intraclass correlation coefficient; mAP@ $k$  = mean average precision at  $k$ ; MEN = meningioma; PA = pilocytic astrocytoma; PCA = principal component analysis; PHATE = potential of heat-diffusion for affinity-based trajectory embedding; t-SNE = t-distributed stochastic neighbor embedding; TICE = T1 contrast-enhanced; UMAP = uniform manifold approximation and projection; WHO = World Health Organization

Existing neuroradiology reference materials largely feature classic imaging presentations of disease entities. Recent surveys reveal that the websites Radiopaedia, StatDx, and UpToDate are particularly popular among radiology trainees<sup>1</sup> and attendings.<sup>2</sup> However, these materials mainly offer curated textbook-style

images and lack the full spectra of disease phenotypes. They also require prior knowledge of relevant search terms, rendering them ill-suited for more inexperienced users.

Though common reference materials include a limited scope of imaging features, institutional databases of historical medical imaging data can address this issue. Technological advancements

Received August 2, 2024; accepted after revision December 11.

From the University of Leipzig (M.v.R., K.W., K.-T.H., S.A.), Leipzig, Germany; Departments of Radiology and Biomedical Imaging (S.C., A.A.), and Therapeutic Radiology (S.A.), Center for Outcomes Research and Evaluation (S.C.), and Center for Translational Imaging Analysis and Machine Learning, Department of Radiology and Biomedical Imaging (T.Z.), Yale School of Medicine, New Haven, Connecticut; Department of Radiology (A.A.), Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; Department of Radiology (N.M., M.S.A.), Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; Department of Neurosurgery (J.L.), Heinrich-Heine-University, Duesseldorf, Germany; Department of Diagnostic and Interventional Radiology, Medical Faculty (N.T.), University Dusseldorf, Dusseldorf, Germany; University of Duisburg-Essen (L.J.), Essen, Germany; DKFZ Division of Translational Neuro-oncology at the WTZ, German Cancer Consortium, DTK Partner Site (L.J.), University Hospital Essen, Essen, Germany; University of Ulm (S.M.), Ulm, Germany; and Visage Imaging, Inc. (M.L.), San Diego, California.

Marc von Reppert and Saahil Chadha contributed equally.

A complete list of the authors in the IBSR Consortium appears at the end of this article.

This work was supported in part by a KL2 TR00186 grant from the National Center for Advancing Translational Sciences.

Please address correspondence to Mariam S. Aboian, Department of Radiology, Children's Hospital of Philadelphia, 3500 Civic Center Blvd, Philadelphia, PA 19104; e-mail: aboianm@chop.edu; @MariamAboian

 Indicates article with supplemental data.

 Indicates article that contains code.

<http://dx.doi.org/10.3174/ajnr.A8805>

## SUMMARY

**PREVIOUS LITERATURE:** Previous studies attempting image-based search for brain tumor MRI retrieval faced practical limitations. Techniques like visual bag of words models, image fusion, and transfer learning were leveraged but these earlier algorithms lacked practicality for clinical use as they required manual tumor delineation by expert neuroradiologists. The need for efficient, automatic methods to streamline image-based search has been emphasized, because reliance on traditional expert-driven segmentation impedes broad clinical application. Furthermore, the utility of retrieved search results was not validated by clinicians. There has been a gap in addressing brain tumor subtype classification with greater precision in retrieval.

**KEY FINDINGS:** This study introduces an algorithm combining deep learning–based automatic tumor segmentation with dimensionality reduction for image-based search. The method outperformed prior algorithms, achieving high retrieval scores across primary brain tumor entities. Its strong clinical validation offers promising potential for neuroradiology education and decision-making.

**KNOWLEDGE ADVANCEMENT:** This research advances knowledge by providing an efficient, automated method for image-based search, optimized for clinical use. Unlike previous methods, it eliminates the need for manual segmentation while demonstrating high retrieval accuracy. Its incorporation of radiomic feature clustering further enhances diagnostic capabilities, improving the integration of artificial intelligence in neuroradiology practice.

have led health care centers to a 10-fold increase in image acquisition between 1999 and 2010.<sup>3</sup> These databases encompass diverse imaging phenotypes with associated histopathologic diagnoses that are established and verified during the clinical care of patients, but most are currently underutilized for education and clinical decision-making.<sup>4</sup> Text-based search engines and machine learning methods are 2 popular methods for leveraging radiologic imaging archives, but both face concerns. Text-based search engines<sup>5</sup> may have more diversity in imaging pathology than classic reference materials, but they require well-crafted search terms. Such engines are familiar to most Internet users, but despite their apparent unicity, these engines often leverage controlled vocabularies, and standardizing free-text fields is complex and imperfect.<sup>6</sup> Moreover, like Radiopaedia and StatDx, text-based search methods rely on a user's skills in crafting search terms. Machine learning offers a promising avenue to leverage archives; however, it faces interpretability and generalizability concerns. Despite the high performance of some primary brain tumor classifiers,<sup>7,8</sup> these models often lack explanations for diagnoses, inducing some apprehension among physicians about their integration at the point-of-care.<sup>9</sup> In addition, many models suffer from a significant drop in performance on out-of-distribution validation data sets from patient populations or health care institutions on which the models were not originally trained.<sup>10</sup>

There is a critical need for a method to retrieve historical medical images based on visual similarity to an image query. Similar methods have commonly been applied to face detection algorithms and other computer vision tasks outside of medicine.<sup>11</sup> Also, previous work in this area<sup>12-15</sup> has been limited in clinical utility by requiring manual tumor segmentation. Here, we introduce an image-based search algorithm that utilizes a convolutional neural network (CNN) for semiautomatic tumor segmentation, radiomics for feature extraction, and dimensionality reduction for clustering. The algorithm automatically retrieves similar reference cases based on extracted query image features without any text input. This image-based search approach is evaluated on an institutional data set of heterogeneous primary brain tumors, while comparing dimensionality reduction techniques. Unlike previous image-based search algorithms, which grouped gliomas together,<sup>12,13</sup> this study is the first to separate glioma subtypes.<sup>16</sup> Performance is

validated by using established retrieval metrics and a visual similarity assessment by 4 board-certified neuroradiologists.

## MATERIALS AND METHODS

### Inclusion Criteria

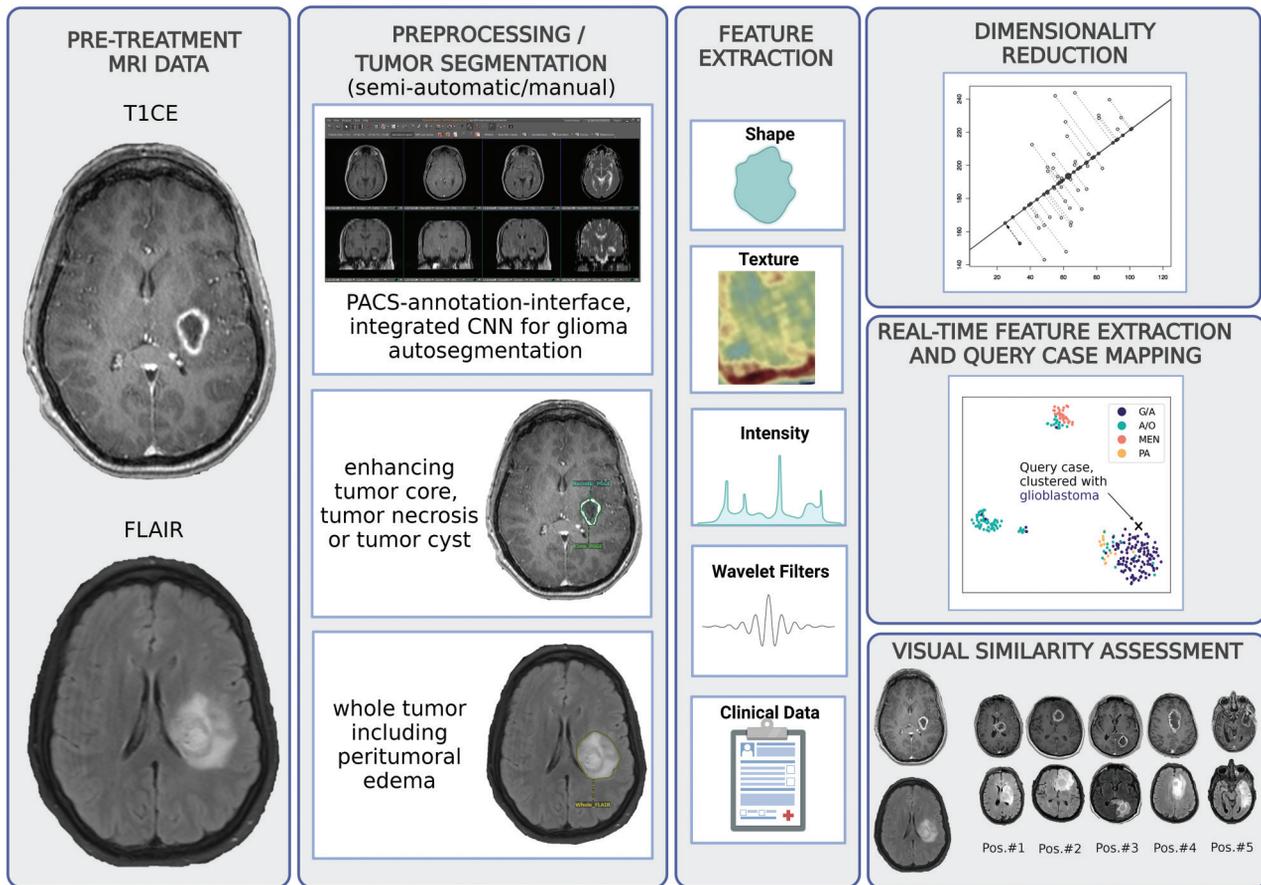
Institutional review board approval was obtained, and a STROBE checklist was used for this retrospective observational study. Initially 1033 consecutive patients treated for primary brain tumors at the study institution between January 2000 and December 2021 were initially considered. Only patients who underwent surgical and/or radiotherapeutic treatment for histopathologically confirmed World Health Organization (WHO) CNS<sup>16</sup> glioblastoma and astrocytoma grade 4 (G/A), pilocytic astrocytoma (PA), astrocytoma and oligodendroglioma grades 2–3 (A/O), and meningioma (MEN) were included. Patient age was not limited. Pretreatment MRIs from internal and external facilities were considered along with those from a pediatric low-grade glioma clinical trial (ClinicalTrials.gov: NCT01734512, PNOC001). Only MRI scans with T1-weighted, T1 contrast-enhanced (T1CE), and FLAIR sequences were used. Patient scans with motion artifact or without hyperintensity on FLAIR were excluded. Overall, 295 patients were included.

### Tumor Segmentation

Three VOIs were segmented: whole tumor on FLAIR as well as enhancing core and necrotic/cystic portions on T1CE. A pretrained U-Net CNN-based automatic segmentation algorithm was utilized for adult-type diffuse gliomas<sup>17</sup> and manually corrected, as necessary. Pediatric-type gliomas and MENs were manually segmented. Segmentations were checked by a neuroradiologist with 7 years of experience (M.S.A).

### Preprocessing and Radiomic Feature Extraction

PyRadiomics Version 3.0.1 was used for image preprocessing and feature extraction.<sup>18</sup> Preprocessing included intensity Z-normalization, 3-mm isotropic voxel resampling, and image discretization. Default MRI parameters were used for feature extraction, which excluded sum average, maximal correlation coefficient, and neighboring gray tone difference matrix features due to



**FIG 1.** PACS-integrated workflow for volume of interest segmentation, feature extraction, dimensionality reduction and clustering of query case. The inputs to our workflow are FLAIR and T1CE images. Semiautomated segmentation was performed within our PACS to segment the whole, enhancing, and necrotic/cystic components of the tumor. Then, 788 radiomic features were extracted from the segmented tumor components. These extracted radiomic features additionally underwent wavelet transformation and were combined with clinical data, including age and sex. Dimensionality reduction was subsequently performed on these features. Reference cases were then projected into dimensionality-reduced space, clustering similar cases together. Finally, for a query case, the nearest reference cases in the dimensionality-reduced space were identified.

redundancy. Fourteen shape-based features were extracted, and 8 high- and low-pass wavelet filters were applied to 18 first-order and 68 gray-level matrix features, resulting in 788 extracted features from each of the 2 T1CE VOIs and 1 FLAIR VOI. The 2364 features in total were utilized together for reference case retrieval.

**Brain Tumor Clustering and Case Retrieval.** Cases were randomly split into reference (85%) and query (15%) subsets. Reference cases' radiomic features were clustered onto the first  $m$  component vectors from dimensionality reduction. Four dimensionality reduction techniques were tested: principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), uniform manifold approximation and projection (UMAP), and potential of heat-diffusion for affinity-based trajectory embedding (PHATE).<sup>19-22</sup> Two to 10 included components vectors ( $m$ ) were tested. Each query case's radiomic features were mapped onto the reduced feature space, and its most similar reference cases were retrieved through a nearest neighbor search. Figure 1 displays a workflow summary.

#### Assessment of Retrieval Performance

Retrieval performance was evaluated by using mean average precision at  $k$  (mAP@ $k$ ), a well-established measurement of the

relevance of retrieved results to a search query<sup>11</sup> previously used in other medical image retrieval studies.<sup>12,13</sup>

Precision at  $k$  represents the proportion of accurate matches among the top  $k$  retrieved cases for a query. For  $r$  accurate matches among the top  $k$  retrieved cases:

$$\text{precision at } k = \frac{r}{k}$$

For  $n$  query cases each with at least  $k$  retrieved reference cases, mAP@ $k$  is defined as:

$$\text{mAP@}k = \frac{1}{n} \sum_{i=1}^n \frac{1}{k} \sum_{j=1}^k \text{precision at } j \text{ for query } i$$

In this study, mAP@5 was the principal metric, as users are less likely to make use of search results beyond the top 5. The technique and number of included components for dimensionality reduction were optimized by maximizing mAP@5.

#### Visual Similarity Assessment

Four expert neuroradiologists (M.S.A, F.M., C.K., and I.I.) with a minimum of 7 years of experience independently assessed visual

## Definition of the 5-point visual similarity scoring system

Score	Definition
5: Very similar	The retrieved case is almost identical to the query case.
4: Similar	The retrieved case is almost identical, though there are differences in imaging features. However, it is subjectively classified to be the same disease.
3: Undecided	Some of the imaging features are similar, while some are different. Radiologically, the likelihood of the same disease is approximately one-half.
2: Dissimilar	The retrieved case has some similarities to the search case, but still looks different.
1: Very dissimilar	The retrieved case is very different from the search case and appears to be a different tumor etiology.

similarity. They compared each of the top 5 retrieved reference MRI lesions to the corresponding query case on a 5-point scale (Table) as introduced previously.<sup>23</sup> Figure 2 shows representative images from this assessment.

### Statistical Analysis

Interrater reliability was assessed by using intraclass correlation coefficient (ICC; 3, k) with Pingouin Version 0.5.3.<sup>24</sup> Values were interpreted as previously reported: poor, <0.5; moderate, 0.5–0.75; good, 0.75–0.9; and excellent, >0.9.<sup>25</sup> A pair-wise *t* test was also performed to compare mAP scores between dimensionality reduction techniques. *P* < .05 was considered statistically significant.

### Data Availability

Data used in this study are available from the authors upon request. Algorithm and analysis code are available through a GitHub repository ([www.github.com/ImagineQuant/radiomics\\_clustering](http://www.github.com/ImagineQuant/radiomics_clustering)).

## RESULTS

### Patient Demographics and Data Split

The Supplemental Data show baseline demographic, clinical, and imaging data by tumor type. Of the 295 patients, 135 (46%) had G/A; A/O, 100 (34%); MEN, 40 (14%); PA, 20 (7%). Moreover, 103 patients' scans (35%) were acquired externally, and the remaining 192 patients' scans (65%) were acquired at the study institution. The reference subset contained 250 cases, and the query subset contained 45 cases.

### Retrieval Performance

A visual representation of reference case clustering by PCA, t-SNE, UMAP, and PHATE can be found in Fig 3A. Maximum mAP@5 values for each dimensionality reduction technique were calculated across every number of included components (Fig 4, Supplemental Data). t-SNE achieved the best performance across tumor types, though it was not statistically significantly better than other techniques (*P* = .18 [PHATE], 0.49 [UMAP], 0.89 [PCA]; Supplemental Data). PCA and UMAP achieved comparable results to t-SNE for G/A, A/O, and MEN, but results differed for PA. PCA achieved a maximum mAP@5 of 73% for PA, and UMAP achieved 68%, compared with 78% for t-SNE. Across tumor types, the poorest performance was observed for PHATE.

Once t-SNE was identified as the highest-performing dimensionality reduction technique, the optimal number of included components was determined by comparing mAP@5 values by tumor type as the number of included components varied from 2–10 (Fig 3B). Including additional t-SNE components beyond 6 did not result in performance improvement. Thus, the final

retrieval model used in qualitative visual similarity scoring incorporated the top 6 t-SNE components.

### Visual Similarity Assessment by 4 Neuroradiologists

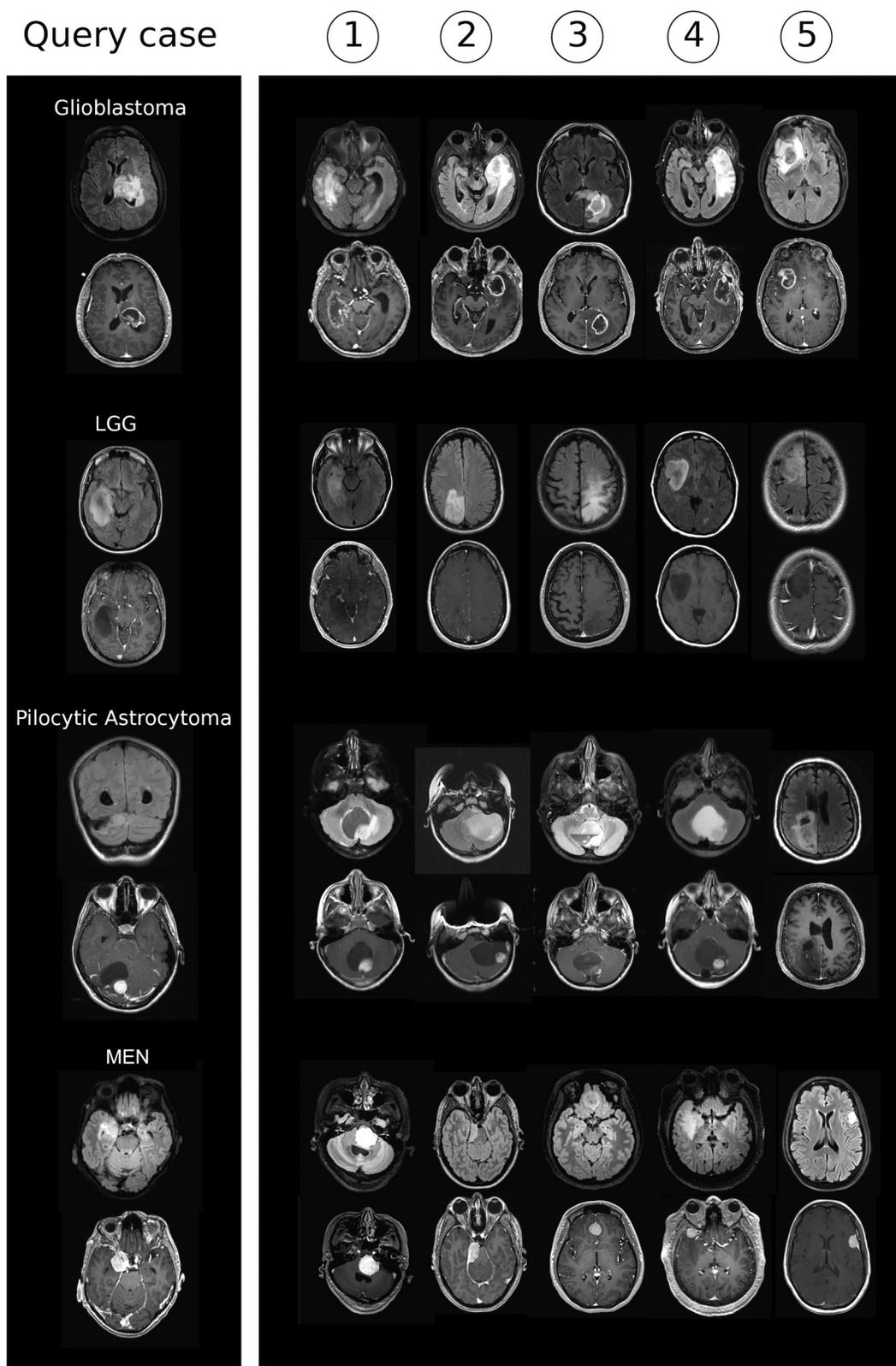
The top-ranked MRI search results demonstrated a mean visual similarity score of  $4.27 \pm 0.82$  overall, with 32% 'similar' and 51% 'very similar' (Table). This score slightly diminished down the rankings, with a mean visual similarity score of  $4.11 \pm 0.96$  for the top 5 results, with 28% 'similar' and 48% 'very similar' (Fig 5, Supplemental Data). The interrater reliability of the overall data demonstrated a good level of agreement, with an ICC (3, k) of 0.78 (95% CI, 0.65–0.87, *P* < .001), providing a good level of reliability between raters. Overall, the top 5 retrieved reference lesions demonstrated high visual similarity with their corresponding query lesions.

Analysis across the 4 tumor groups exhibited high agreement levels for G/A and PA, each demonstrating an ICC (3, k) of 0.95 (G/A: *P* < .001; 95% CI, 0.90–0.98; PA: *P* < .001; 95% CI, 0.84–0.99). A/O tumors had similar ICC (3, k) values of 0.92 (*P* < .001; 95% CI, 0.83–0.96). While there was a moderate level of agreement among neuroradiologists in their similarity assessments of MENs (ICC [3, k], 0.55), this level of agreement was not statistically significant (*P* = .05; 95% CI, –0.19 to 0.87).

## DISCUSSION

Many radiologists and trainees rely on reference materials that focus on classic imaging features and require user expertise; the full spectrum of brain tumor imaging appearance is only learned over years of clinical practice. To address these challenges, we proposed an image-based search technique to retrieve primary brain tumor reference cases based on imaging appearance. This method has potential applications in differential diagnosis formulation, clinical and research database development, characterization of brain tumor subtype imaging features, and educational initiatives.

We demonstrated that dimensionality reduction and radiomic feature clustering is a promising method for performing image-based search of radiologic databases. t-SNE was the best-performing dimensionality reduction technique, achieving near-perfect mAP@5 scores for G/A and MENs. Performance on PA was comparatively lower at 78%, attributable to fewer cases and overlapping contrast enhancement features between PAs and glioblastomas/MENs. Queries for A/O also had a comparatively lower mAP@5 score of 88%, likely due to overlapping imaging features on FLAIR. Expert neuroradiologists' visual similarity assessment demonstrated overall high performance of our approach, with a mean visual similarity score of  $4.11 \pm 0.99$  of 5 (76% 'similar' or

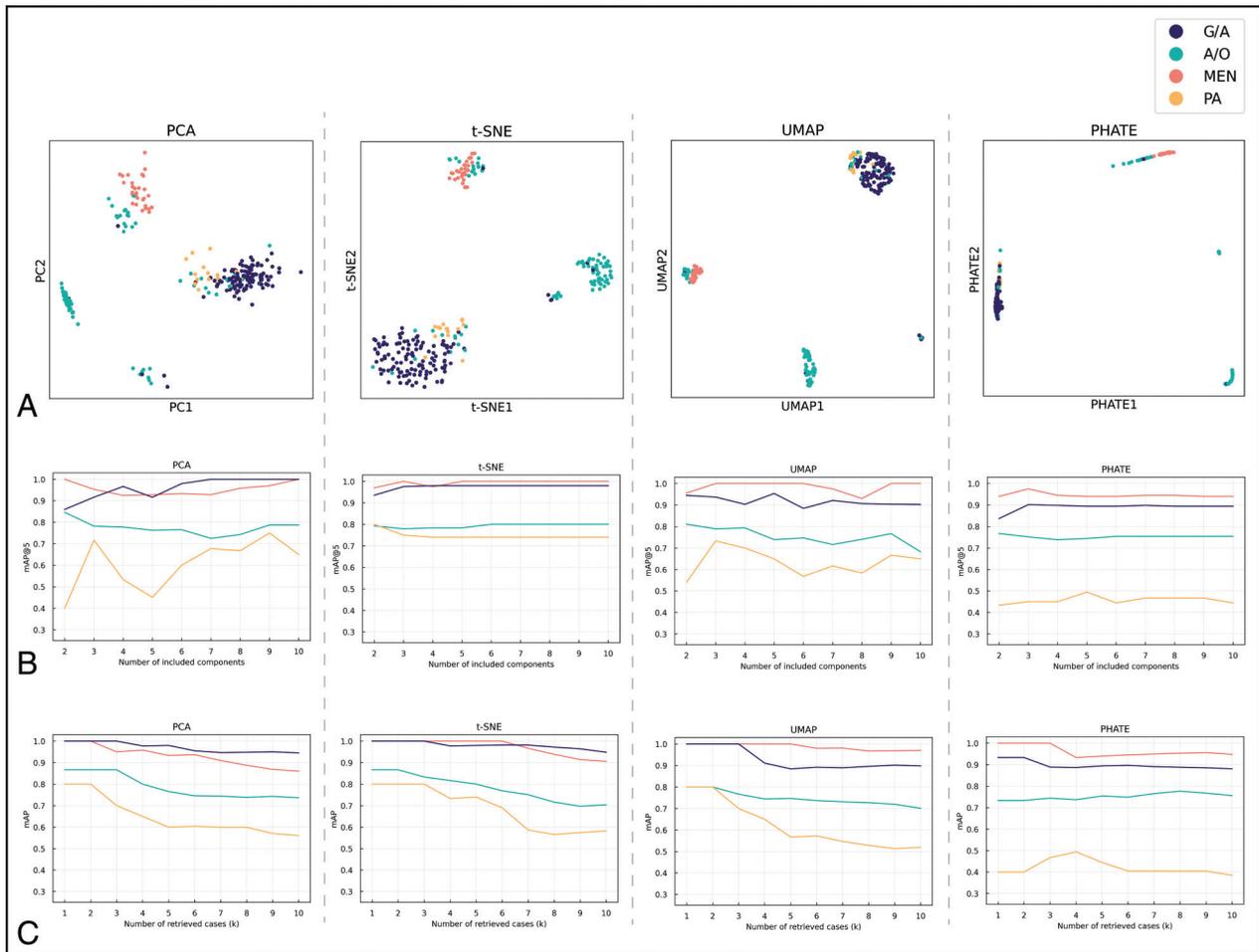


**FIG 2.** Axial FLAIR and T1CE MRIs of example query cases with corresponding 5 top-retrieved results. LGG indicates low-grade glioma.

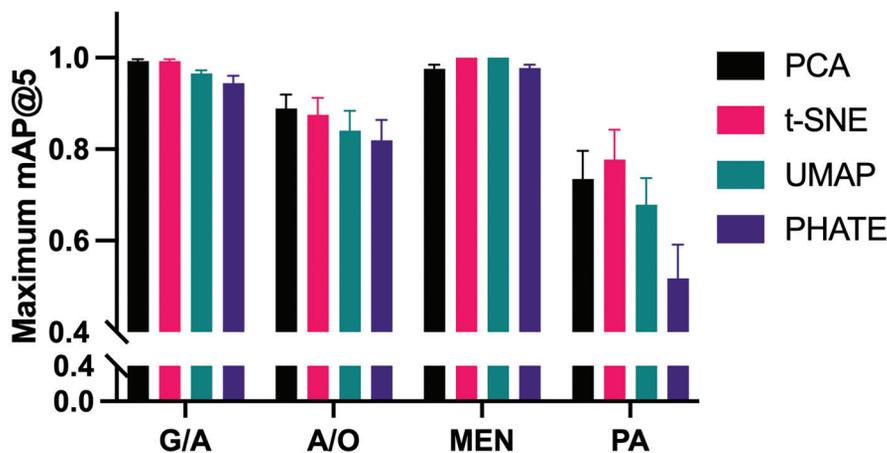
‘very similar’) for the top 5 retrieved results. This highlights the promising utility of image-based search in clinical practice.

Prior studies<sup>12-15</sup> attempted to perform image-based search for retrieval of primary brain tumor MRIs but were ill-suited for clinical

implementation due to practical constraints. Though in practice our algorithm retrieves all reference cases in a ranked list, we utilized the mAP@5 metric to compare performance with this previous work. Retrieval performance of these earlier algorithms was fair;



**FIG 3.** Retrieval case clustering and performance. *A*, Retrieval case clustering of the first 2 components. *B*, mAP@5 by varying number of components. *C*, mAP@*k* for 6 included components and varying number of retrieved cases (*k*).



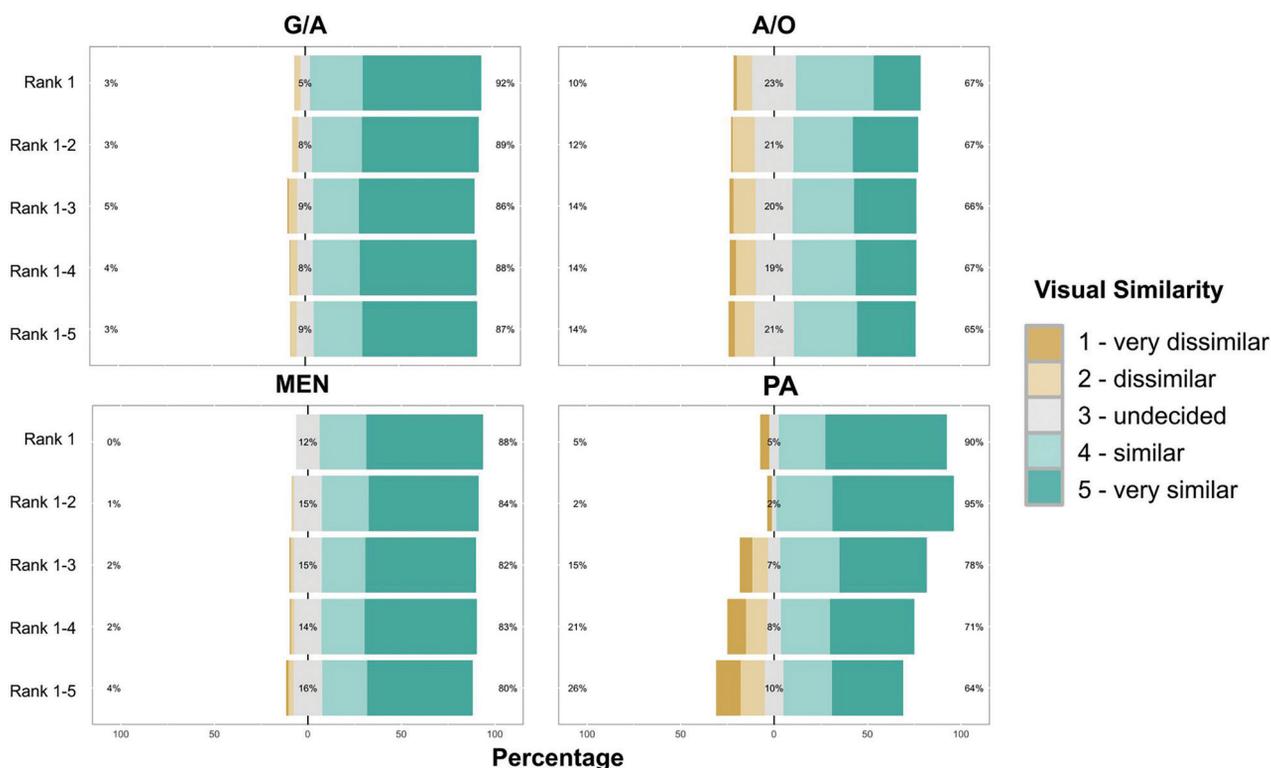
**FIG 4.** Maximum mean average precision of top 5 retrieved cases by dimensionality reduction technique and tumor type. Error bars represent standard errors of the mean.

2 algorithms used visual bag of words models,<sup>12,13</sup> and others used image fusion<sup>14</sup> and transfer learning.<sup>15</sup> These achieved overall maximum mAP@5 values of 92% for MENs and 98% for gliomas. Our image-based search algorithm outperformed all these previous algorithms for MEN retrieval (mAP@5, 100%),

and average weighted performance on glioma subtypes was comparable (mean mAP@5, 93%). Crucially, however, all the earlier-mentioned algorithms required manual tumor segmentation from expert neuroradiologists, a known time-consuming and subjective process.<sup>26</sup> Clinical implementation of such algorithms would be limited by this bottleneck. We depart from these earlier, less practical algorithms by incorporating a CNN-based model for automatic tumor segmentation. Together with dimensionality reduction to minimize computational load, our image-based search algorithm is directly optimized for clinical utility. Furthermore, unlike earlier work, we

clinically validate our search algorithm by using expert neuro-radiologist assessment.

There are a few limitations to this study. The introduced algorithm cannot distinguish visually similar primary brain tumor subtypes due to their grouping (eg, low-grade astrocytoma/oligodendroglioma),



**FIG 5.** Likert lesion visual similarity scores of top 1–5 retrieved cases by tumor type. Visual similarity scoring was conducted on a scale of 1 (very dissimilar) to 5 (very similar) based on the appearance of each retrieved reference lesion to the query lesion. The similarity scores for the top 5 cases were iteratively pooled in descending order to represent use in clinical practice. The percentage of cases rated ‘very dissimilar’ or ‘dissimilar’ are along the left axes, the percentage rated ‘similar’ or ‘very similar’ are along the right axes, and the percentage rated ‘undecided’ are along the center axis of each plot.

a common challenge in image-based search focused on maximizing signal-to-noise ratios. Notably, to the authors’ knowledge, this is the first study to perform image-based search of primary brain tumors while separating glioma subtypes. Future work will provide imaging characteristics to be used in our proposed pipeline to distinguish between tumors such as grade 2 and grade 3 *IDH*-mutant astrocytomas. Additionally, our algorithm does not account for anatomic location, yet retrieval performance remains strong regardless. Moreover, visual similarity assessment involved only 4 expert raters, but high ICC values indicate reliability of these results. Finally, although this study utilized a large, heterogeneous data set, multi-institutional testing will be used to further validate the proposed algorithm and determine the optimal degree of generalization and customization for scaling.

Through image-based search, neuroradiologists and trainees can identify reference cases that share visual similarities with their patients’ scans. The performance of this approach has been validated through both quantitative and qualitative assessments, and it has shown promise for clinical application. Image-based search can provide valuable guidance in neuroradiology education, clinical decision-making, and research.

## CONCLUSIONS

We propose a method to search historical MR images of primary brain tumors and retrieve reference cases highly similar to presented

query cases. Evaluation through tumor type comparison and visual similarity scoring by expert neuroradiologists corroborated the high performance of this approach. If translated into a clinical setting, image-based search will offer a reference material that encompasses the full diversity of imaging pathology.

## ACKNOWLEDGMENTS

We would like to thank Dr. XP who was instrumental in project conception and design. This research was supported by the Richard K. Gershon Endowed Medical Student Research Fellowship and Yale School of Medicine Fellowship for Medical Student Research.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## IBSR CONSORTIUM

Manpreet Kaur, Ludwig Maximilian University, Munich, Germany; Divya Ramakrishnan, Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, Connecticut; Khaled Bousabarah, Visage Imaging GmbH, Berlin, Germany; Nicolas Linder, University of Leipzig, Leipzig, Germany; Sarah M. Jacobs, Center for Image Sciences, University Medical Center Utrecht, Utrecht, the Netherlands; Francesca Conti, Sapienza Università di Roma, Rome, Italy; Gabriel Cassinelli Petersen, University of Göttingen, Göttingen, Germany; Ujjwal

Baid, Indiana University School of Medicine, Indianapolis, Indiana; Ichiro Ikuta, Department of Radiology, Mayo Clinic Comprehensive Cancer Center, Phoenix, Arizona; Fatima Memon, Carolina Radiology, Myrtle Beach, South Carolina; Claudia Kirsch, Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, Connecticut; Melissa Davis, Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, Connecticut; Spyridon Bakas, Indiana University School of Medicine, Indianapolis, Indiana.

## REFERENCES

1. Derakhshani A, Ding J, Vijayasarithi A. **On-call radiology 2020: where trainees look for help in a high stakes and time sensitive environment.** *Clin Imaging* 2021;77:219–23 CrossRef Medline
2. Niederhauser BD, Liaw K, McDonald RJ, et al. **Pick up a book or “Google It?” A survey of radiologist and trainee-preferred references and resources.** *J Digit Imaging* 2014;27:26–32 CrossRef
3. McDonald RJ, Schwartz KM, Eckel LJ, et al. **The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload.** *Acad Radiol* 2015;22:1191–98 CrossRef Medline
4. Akgül CB, Rubin DL, Napel S, et al. **Content-based image retrieval in radiology: current status and future directions.** *J Digit Imaging* 2011;24:208–22 CrossRef Medline
5. Demner-Fushman D, Kohli MD, Rosenman MB, et al. **Preparing a collection of radiology examinations for distribution and retrieval.** *J Am Med Inform Assoc* 2016;23:304–10 CrossRef Medline
6. Sotomayor CG, Mendoza M, Castañeda V, et al. **Content-based medical image retrieval and intelligent interactive visual browser for medical education, research and care.** *Diagnostics (Basel)* 2021; 11:1470 CrossRef Medline
7. Deepak S, Ameer PM. **Brain tumor classification using deep CNN features via transfer learning.** *Comput Biol Med* 2019;111:103345 CrossRef Medline
8. Kokkalla S, Kakarla J, Venkateswarlu IB, et al. **Three-class brain tumor classification using deep dense inception residual network.** *Soft Comput* 2021;25:8721–29 CrossRef Medline
9. Geis JR, Brady AP, Wu CC, et al. **Ethics of artificial intelligence in radiology: summary of the Joint European and North American Multisociety Statement.** *Radiology* 2019;293:436–40 CrossRef Medline
10. Eche T, Schwartz LH, Mokrane F-Z, et al. **Toward generalizability in the deployment of artificial intelligence in radiology: role of computation stress testing to overcome underspecification.** *Radiology Artif Intell* 2021;3:e210097 CrossRef Medline
11. Latif A, Rasheed A, Sajid U, et al. **Content-based image retrieval and feature extraction: a comprehensive review.** *Mathematical Problems in Engineering* 2019;2019:1–21 CrossRef
12. Huang M, Yang W, Wu Y, et al. **Content-based image retrieval using spatial layout information in brain tumor T1-weighted contrast-enhanced MR images.** *PLoS One* 2014;9:e102754
13. Huang M, Yang W, Yu M, et al. **Retrieval of brain tumors with region-specific bag-of-visual-words representations in contrast-enhanced MRI images.** *Comput Math Methods Med* 2012;2012: e280538 CrossRef Medline
14. Dube S, El-Saden S, Cloughesy TF, et al. **Content based image retrieval for MR image studies of brain tumors.** *Conf Proc IEEE Eng Med Biol Soc* 2006;2006:3337–40 CrossRef Medline
15. Swati ZNK, Zhao Q, Kabir M, et al. **Content-based brain tumor retrieval for MR images using transfer learning.** *IEEE Access* 2019; 7:17809–22 CrossRef
16. Louis DN, Perry A, Wesseling P, et al. **The 2021 WHO Classification of Tumors of the Central Nervous System: a summary.** *Neuro Oncol* 2021;23:1231–51 CrossRef Medline
17. Aboian M, Bousabarah K, Kazarian E, et al. **Development of a workflow efficient PACS based automated brain tumor segmentation and radiomic feature extraction for clinical implementation (N2.003).** *Neurology* 2022;98:3146 CrossRef
18. van Griethuysen JJM, Fedorov A, Parmar C, et al. **Computational radiomics system to decode the radiographic phenotype.** *Cancer Res* 2017;77:e104–07 CrossRef Medline
19. Pearson K. **LIII. On lines and planes of closest fit to systems of points in space.** *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 1901;2:559–72 CrossRef
20. Maaten L, van der Hinton G. **Visualizing data using t-SNE.** *J Mach Learn Res* 2008;9:2579–605.
21. McInnes L, Healy J, Melville J. **Umap: uniform manifold approximation and projection for dimension reduction.** arXiv.org 2018 <https://arxiv.org/abs/1802.03426> (accessed December 10, 2023)
22. Moon KR, van Dijk D, Wang Z, et al. **Visualizing structure and transitions in high-dimensional biological data.** *Nat Biotechnol* 2019;37:1482–92 CrossRef Medline
23. Hwang HJ, Seo JB, Lee SM, et al. **Content-based image retrieval of chest CT with convolutional neural network for diffuse interstitial lung disease: performance assessment in three major idiopathic interstitial pneumonias.** *Korean J Radiol* 2021;22:281–90 CrossRef Medline
24. Vallat R. **Pingouin: statistics in Python.** *JOSS* 2018;3:1026 CrossRef
25. Koo TK, Li MY. **A guideline of selecting and reporting intraclass correlation coefficients for reliability research.** *J Chiropr Med* 2016; 15:155–63 CrossRef Medline
26. Multi-Institutional Target Delineation in Oncology Group. **Human-computer interaction in radiotherapy target volume delineation: a prospective, multi-institutional comparison of user input devices.** *J Digit Imaging* 2011;24:794–803 CrossRef Medline

# The Value of Quantitative Susceptibility Mapping and Morphometry in the Differential Diagnosis of Parkinsonism

Yi Li, Tingting Yuan, Lulu Gao, Wei Sun, Xiaoxiao Du, Zhihui Sun, Kangli Fan, Ruqing Qiu, and Ying Zhang



## ABSTRACT

**BACKGROUND AND PURPOSE:** Differentiating Parkinson disease (PD) from atypical parkinsonism syndrome (APS), including multiple system atrophy (MSA) and progressive supranuclear palsy (PSP), is challenging, and there is no gold standard. Integrating quantitative susceptibility mapping (QSM) and morphometry can help differentiate PD from APS and improve the internal diagnosis of APS.

**MATERIALS AND METHODS:** In this retrospective study, we enrolled 55 patients with PD, 17 with MSA-parkinsonian type (MSA-P), 15 with MSA-cerebellar type (MSA-C), and 14 with PSP. Thirty-three age-matched healthy subjects served as controls. All subjects underwent QSM imaging and 3D TIWI with manual quantification of ROI and morphometry. ROIs were selected in the basal ganglia and brainstem nuclei, such as the putamen (Pu), globus pallidus (GP), and red nucleus (RN). Morphometry included MR Parkinson disease index (MRPI), the midbrain area-pons area ratio (M/P), and the ratio of the vertical line of the long axis of the midbrain and pons (Ratio). Differential variables between groups were extracted and a binary logistic regression was established to differentiate the differential diagnoses of PD and APS and diseases within APS. The diagnostic value was assessed using the area under the curve (AUC), sensitivity, and specificity.

**RESULTS:** The combination of Pu and GP performed best when used to distinguish PD from MSA-P, with an AUC of 0.800 (95% CI: 0.664–0.936). The AUC was optimal when MRPI and M/P were combined to distinguish PD from MSA-C at 0.823 (95% CI: 0.686–0.960). Ratio alone performed best in differentiating PD from PSP, with an AUC of 0.848 (95% CI: 0.711–0.985). The AUC for Ratio alone in distinguishing MSA-P from PSP was 0.871 (95% CI: 0.738–1.0). The AUC when using only M/P to distinguish MSA-C from PSP was 0.931 (95% CI: 0.845–1.0). QSM and morphometry each offer distinct advantages in the differential diagnosis among the aforementioned groups. The combination of QSM and morphometry provided the highest diagnostic value in differentiating PD from APS, highlighting the significance of integrating these 2 imaging techniques for enhanced diagnostic precision in clinical practice. The best indicators described above showed equally high differential diagnostic values in patients with a disease duration of  $\leq 3$  years.

**CONCLUSIONS:** QSM and morphometry will improve the differential diagnosis between PD and APS, as well as improve the internal diagnosis of APS.

**ABBREVIATIONS:** APS = atypical parkinsonism syndrome; AUC = area under the curve; DN = dental nucleus; GP = globus pallidus; H-Y = Hoehn and Yahr Scale; ICC = intraclass correlation coefficient; M/P = midbrain area-pons area ratio; MRPI = magnetic resonance parkinsonism index; MSA = multiple system atrophy; MSA-C = multiple system atrophy cerebellar subtype; MSA-P = multiple system atrophy parkinsonian subtype; PD = Parkinson disease; PSP = progressive supranuclear palsy; Pu = Putamen; QSM = quantitative susceptibility mapping; Ratio = the ratio of the vertical line of the long axis of the midbrain and pons; RN = red nucleus; SN = substantia nigra; SNc = substantia nigra compacta; Th = thalamus; UPDRS-III = Unified Parkinson Disease Rating Scale Part III

**P**arkinson disease (PD) is a common neurodegenerative disease. Its basic pathologic changes are degeneration and loss

of dopaminergic neurons in the substantia nigra and formation of Lewy bodies in the cytoplasm of surviving neurons. The loss of dopaminergic neurons causes a decrease in the content of dopamine input to the striatum, which leads to typical extrapyramidal movement disorders, such as bradykinesia, rigidity, and static tremor. Multiple system atrophy (MSA) and progressive

Received July 19, 2024; accepted after revision January 16, 2025.

From the Departments of Neurology (Y.L., W.S., X.D., Z.S., K.F., R.Q., Y.Z.) and Radiology (T.Y.), the First Hospital of Jilin University, Changchun, Jilin Province, China; Department of Neurology (Y.L.), Fujian Medical University Union Hospital, Fuzhou, Fujian Province, China; and Department of Neurology (L.G.), the First People's Hospital of Jining, Jining, Shandong Province, China.

Yi Li and Tingting Yuan contributed equally to this article.

This research was funded by grants from the Natural Science Foundation of Jilin Province (NO. YDZJ202201ZYT5116), the Jilin Province Department of Finance Project (NO. JLSWSRCZX2020-0011), and the Jilin Province Health and Technology Innovation Project (NO. 2018J045) for their generous support to Y.Z.

Please address correspondence to Ying Zhang, MD, Department of Neurology, the First Hospital of Jilin University, No. 1 Xinmin St, Changchun, 130000, China; e-mail: zhang\_ying99@jlu.edu.cn

Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8665>

## SUMMARY

**PREVIOUS LITERATURE:** Previous studies have shown that QSM and morphometry can detect brain iron deposition and structural changes associated with Parkinsonism. Increased susceptibility values in certain brain regions, notably the substantia nigra and basal ganglia, correlate with PD and APS. Morphometry provides direct measurements of long-term changes in brain structure, which is very useful for monitoring disease progression. Morphometry also reveals the atrophy of the brain, indicating neurodegeneration.

**KEY FINDINGS:** QSM and morphometry can distinguish PD from APS and the identification of disease within APS by quantifying brain iron and structural alterations.

**KNOWLEDGE ADVANCEMENT:** Combining QSM with morphometry enhances diagnostic value in Parkinsonism, offering new avenues for early detection and differential diagnosis.

supranuclear palsy (PSP) in atypical parkinsonian syndrome (APS) overlap with PD in terms of clinical and pathologic impairments. However, the response to treatment and prognosis of PD and APS are very different,<sup>1-3</sup> which poses a further challenge for the differential diagnosis of PD and APS.

Specific MR signs have been described based on neurodegeneration patterns in PD, MSA, and PSP,<sup>4</sup> but they tend to be less sensitive.<sup>5,6</sup> Though both PD and MSA are synucleinopathies, the degenerative process of MSA primarily involves the dorsolateral portions of the putamen and pons, the middle cerebellar peduncle, and the cerebellum.<sup>7</sup> MSA-parkinsonian subtype (MSA-P) involves atrophy and signal changes of the putamen,<sup>2,8</sup> and MSA-cerebellar subtype (MSA-C) shows mainly atrophy of the pons, middle cerebellar peduncle, and cerebellum.<sup>2,9</sup> However, the degenerative process in PD patients generally does not involve the pons and cerebellum. PSP is a tauopathy with early involvement of the midbrain and superior cerebellar peduncle,<sup>10</sup> and MRI may show atrophy of the midbrain and superior cerebellar peduncle.<sup>8,11,12</sup> Therefore, morphologic changes in the brain structure of patients with PD and APS may be helpful in the differential diagnosis between them. Additionally, damage to the substantia nigra (SN) is common in these disorders, but the neuronal loss in the substantia nigra compacta (SNc) proceeds with different patterns: the lateral part of the ventral layer of SNc is the most affected in PD, the medial part is more affected in PSP.<sup>13</sup>

MRI routine sequences play an important role in the search for secondary causes of Parkinson syndrome but are of limited value in the early diagnosis of neurodegenerative diseases. Satoru et al<sup>14</sup> used 3D-FIESTA imaging to measure the olfactory bulb area, finding it significantly smaller in PD patients with olfactory dysfunction compared with those with APS. However, distinguishing between PD and APS patients without olfactory dysfunction was challenging, and measuring the olfactory bulb area was difficult. Compared with DTI or fMRI, morphometry provides direct measurements of long-term changes in brain structure, which is very useful for monitoring disease progression. While DTI and fMRI contribute to the diagnosis and differential diagnosis of neurodegenerative diseases, their research results often show heterogeneity.

The etiology and pathogenesis of PD are unknown, but dynamic iron imbalance is considered to be a potential pathogenic factor.<sup>15</sup> A correlation between iron deposition in the attenuated part of the

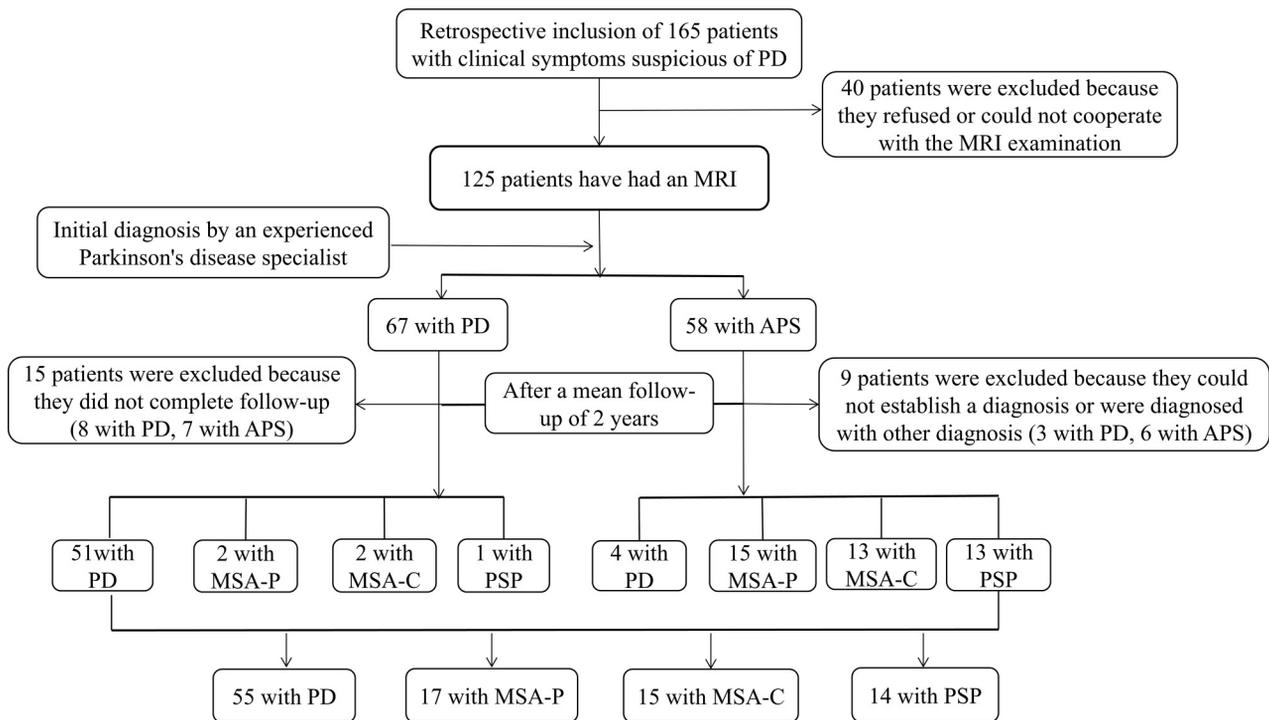
substantia nigra of the midbrain and dopaminergic neuronal damage in patients with PD has been demonstrated.<sup>16,17</sup> In addition, Boelmans et al<sup>18</sup> found that diseases such as MSA and PSP may also have abnormal patterns of brain iron deposition.<sup>19</sup> Thus, identifying the characteristics of iron deposition in neurodegenerative diseases is expected to provide new evidence for diagnosis. MRI sequences are sensitive to tissue susceptibility rates and are commonly used for iron studies, such as SWI and quantitative susceptibility mapping (QSM). Though SWI is highly sensitive in detecting and displaying iron deposits, it cannot quantitatively measure the susceptibility values of tissues, which leads to the inability of SWI to distinguish between PD and APS.<sup>20,21</sup> QSM can quantify iron deposition, and studies have shown that the results of QSM detection are consistent with autopsy studies.<sup>22</sup>

QSM and morphometry are increasingly used in neurodegenerative diseases. However, most studies have focused on the later stages of disease progression. Our study not only included patients across the full course of the disease but also particularly focused on a population of patients diagnosed for 3 years or less. We investigated the combined diagnostic value of QSM and morphometry in differentiating PD from APS and internal differential diagnosis of APS.

## MATERIALS AND METHODS

### Subjects

Retrospective consecutive 165 patients with suspected PD who attended the Movement Disorders Clinic of the First Hospital of Jilin University between December 2021 and June 2023 were included. Patients underwent MRI, including 3D T1WI structural and QSM imaging sequences at baseline visit. Using the Unified Parkinson Disease Rating Scale Part III (UPDRS-III),<sup>23</sup> Hoehn and Yahr Scale (H-Y)<sup>24</sup> to assess motor symptoms in the “off” state at baseline. After a mean follow-up of 2 years, the final diagnosis was made by experienced specialists based on the relevant diagnostic criteria.<sup>1-3</sup> MSA was divided into the MSA-P and MSA-C, and mixed patients were assigned to the MSA-P group. Figure 1 shows the flowchart of patient enrollment. Given the clinical importance of early and accurate diagnosis, a subgroup analysis was also conducted for patients with  $\leq 3$  years of disease duration. This analysis aimed to determine whether the diagnostic methods maintain their performance in early-stage disease, where differentiation between disorders is often more challenging.



**FIG 1.** The flowchart of patient enrollment.

Controls were recruited from the Physical Examination Center of the First Hospital of Jilin University from December 2021 to June 2023, according to age-matching. Subjects were excluded if they had any additional neurologic disorders. The Ethical Review Committee of the First Hospital of Jilin University approved this study. Ethics Approval Number: 23K052-001. All subjects participated in this study voluntarily and signed an informed consent form.

### Imaging Acquisition

MRI was performed on a 3T scanner equipped with an 8-channel head coil. Fast gradient-echo (TFE) was used to obtain high-resolution T1-weighted 3D anatomic brain images (repetition time = 7.0 ms; echo time = 3.2 ms; flip angle = 7°; voxel size = 1 × 1 × 1 mm, slice orientation: sagittal planes). The QSM protocol parameters measured with GRE sequence are as follows (repetition time = 50 ms; echo time/δ TE = 5.6/10.0 ms; flip angle = 20°; voxel size 1 × 1 × 1.5 mm, slice orientation: axial planes).

### Data Processing

The QSM images were postprocessed using the software Matlab R2016a with the STI Suite 3.0 plug-in installed to obtain the magnetic susceptibility parameter map. STI Suite for QSM post-processing includes Laplacian-based phase unwrapping, the iHARPERELLA method for integrated phase unwrapping and background phase removal, and QSM uses the STAR method. The Matlab processed images in nii format and imported them into ITK-SNAP 3.8 software for drawing. Manual measurements of the ROI were performed independently in the best 3 layers of nucleus imaging at 3 axial levels of fixation by a neuroradiologist and a neurology clinician who were unaware of

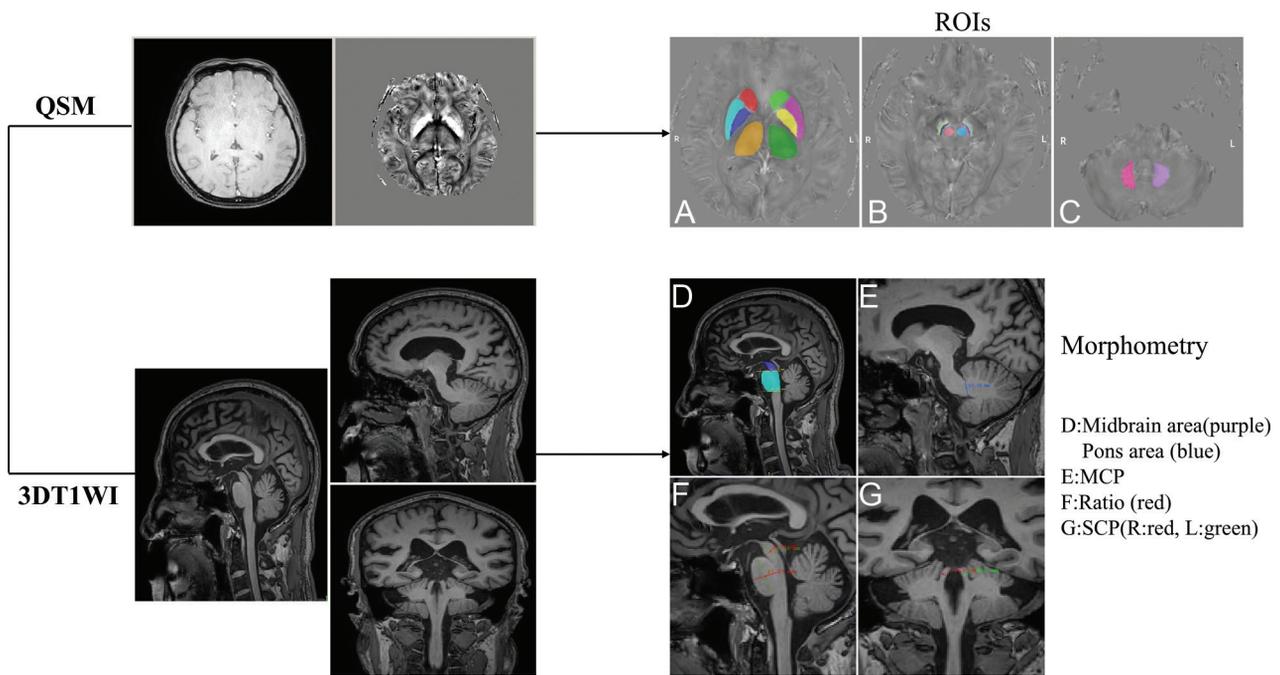
the diagnostic results. Following an established protocol, the SN was divided into the dorsal SNc and the ventral substantia nigra reticulata (SNr).<sup>25</sup> The extra-nigral regions include upstream structures receiving nigral projections, such as caudate nucleus (CN), putamen (Pu), and globus pallidus (GP).<sup>26</sup> The extra-nigral regions are more affected by neurodegeneration in MSA or PSP, such as the red nucleus (RN), thalamus (Th), and dental nucleus (DN). All the above ROIs were measured bilaterally and then averaged, as shown in Fig 2.

Subjects completed the MRI containing the T1WI sequence with no further image postprocessing. Data were subsequently analyzed using the hospital's Picture Archiving and Communication System (PACS) system. According to published methods, 2 doctors manually measured morphometric parameters in the 3D T1WI sagittal and coronal section of MRI, respectively. Morphometric parameters include the midbrain area-pons area ratio (M/P), MR parkinsonism index (MRPI) = (P/M) × (middle cerebellar peduncle width/superior cerebellar peduncle width), the ratio of the vertical line of the long axis of the midbrain and pons (Ratio).<sup>27-29</sup> These morphometric parameters were mainly used to assess the degree of atrophy of the midbrain, pons, and cerebellum in patients with PD and APS. The 2 doctors measured the morphometric parameters 3 times and then averaged them, as shown in Fig 2.

Measurements were performed 3 times by the same doctor in 3 different sessions to assess the intrarater reliability and by a second doctor to assess the interrater reliability. The first measurement is completed within 1 week of the MRI, followed by the last 2 measurements within 6 months.

### Statistical Analyses

The Shapiro-Wilk test was used to test whether the quantitative data were normally distributed. Continuous quantitative data



**FIG 2.** Image processing flow. QSM: The ROIs of QSM include parts A, B, and C. A, The basal ganglia area, including the CN: green (L) and red (R), putamen (Pu): pink (L) and blue (R), GP: yellow (L) and purple (R), Th: dark green (L) and orange (R). B, The midbrain region, including RN: blue (L) and pink (R), compact part of SNc: grass green (L) and soft green (R), SNr: purple (L) and navy blue (R). C, the cerebellar region, including DN: purple (L) and pink (R). 3D TIWI: Morphometrics were measured on TIWI, including parts D, E, F, and G. D, The midbrain area and pons area (MRI midsagittal section): midbrain area (purple) and pons area (blue). E, MCP, middle cerebellar peduncle width (MRI midsagittal section, average of both sides, blue line). F, Ratio, the ratio of the vertical line of the long axis of the midbrain and pons (MRI midsagittal section, ratio of two red lines). G, SCP, superior cerebellar peduncle width (MRI coronal section, average of both sides, red and green lines).

were described by the mean  $\pm$  standard deviation. The  $X^2$  test was applied to analyze sex differences among groups, and it was corrected according to the Bonferroni method. Evaluation with ANOVA, followed by Bonferroni correction and Tamhane's correction test, was performed to compare age, disease duration, UPDRS-III score, and H-Y stage. A one-way ANOVA test was used to assess the differences in the susceptibility values of ROIs and morphometric parameters among groups, followed by evaluation with the Bonferroni correction and Tamhane's correction test for multiple comparisons. Logistic regression was performed to control for the effect of the UPDRS-III score and H-Y stage on imaging variables by using the disease group as the dependent variable and UPDRS-III score, H-Y stage, the susceptibility values of ROIs, and morphometric parameters as independent variables. Binary logistic regression determined the best indicator combination for differentiating 2 diseases (PD versus MSA-P, PD versus MSA-C, PD versus PSP, MSA-P versus PSP, MSA-C versus PSP, PD versus APS). To prevent possible overfitting of the model, we simplified the model by reducing the number of covariates used in the subgroup analysis. Using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve, sensitivity, and specificity to evaluate diagnostic value. The optimal cutoff values were the values that had the highest sum of sensitivity and specificity. The interrater reliability of manual measurements was assessed using intraclass correlation coefficients (ICCs) for the 2 doctors. The study followed the methodology proposed in the STARD checklist.  $P < .05$  was defined as statistically significant. Data analyses were undertaken using IBM SPSS Statistics 27. Plotting was performed using GraphPad Prism 8.

## RESULTS

### Subject Characteristics

We included 55 PD, 17 MSA-P, 15 MSA-C, and 14 PSP patients and 33 healthy controls. The UPDRS-III scores ( $F = 2.777$ ,  $P = .04$ ) and the H-Y stage ( $F = 9.969$ ,  $P < .001$ ) were statistically different among all groups. No difference in sex, age, and disease duration between patients with PD, MSA, and PSP and controls. Demographic and clinical characteristics of patients and controls are summarized in Table 1.

### Intergroup Comparison of QSM and Morphometry

The susceptibility values of all ROIs considered in this study are reported in Supplemental Data; significant differences across patient groups were observed in Pu ( $F = 5.231$ ,  $P < .001$ ), GP ( $F = 2.741$ ,  $P = .04$ ), and red nucleus (RN) ( $F = 2.933$ ,  $P = .02$ ).

The susceptibility values of the Pu in the MSA-P group were higher than those in the Control ( $P = .01$ ), PD ( $P < .001$ ), and PSP ( $P = .02$ ) groups (Fig 3A). The susceptibility values of the GP in the MSA-P group were higher than those in the Control ( $P = .04$ ) and PD ( $P < .001$ ) groups (Fig 3B). However, the susceptibility values of the RN in the PSP group were higher than in the Control ( $P = .02$ ) and PD ( $P = .049$ ) groups (Fig 3C). No differences in susceptibility values for the other ROIs (SNc, SNr, CN, Th, DN) were found between the PD, MSA-P, MSA-C, PSP, and Control groups (Supplemental Data).

There were significant differences in MRPI ( $F = 8.340$ ,  $P < .001$ ), M/P ( $F = 9.457$ ,  $P < .001$ ), and Ratio ( $F = 6.220$ ,  $P < .001$ ) among the groups. The MRPI of the PSP group was higher than the Control ( $P = .007$ ) and MSA-C ( $P < .001$ ) groups. The MRPI

**Table 1: Study cohort**

	Control	PD	MSA-P	MSA-C	Psp	F/X <sup>2</sup>	P value
Number	33	55	17	15	14	—	—
Sex, men/women (N)	13/20	30/25	4/13	8/7	9/5	7.660	.11
Age (years)	58.2 ± 13.1	59.2 ± 9.2	56.5 ± 8.6	54.3 ± 7.4	61.5 ± 5.7	2.380	.07 <sup>b</sup>
Disease duration (years)	—	3.4 ± 2.4	2.8 ± 1.8	2.7 ± 2.0	4.3 ± 3.9	0.953	.43 <sup>b</sup>
UPDRS-III score	—	31.5 ± 21.4	40.6 ± 22.4	41.5 ± 22.9	47.9 ± 19.1	2.777	.04 <sup>a,c</sup>
H-Y stage	—	1.9 ± 0.8	2.2 ± 1.0	3.2 ± 1.2	2.9 ± 1.2	9.969	<.001 <sup>a,c</sup>

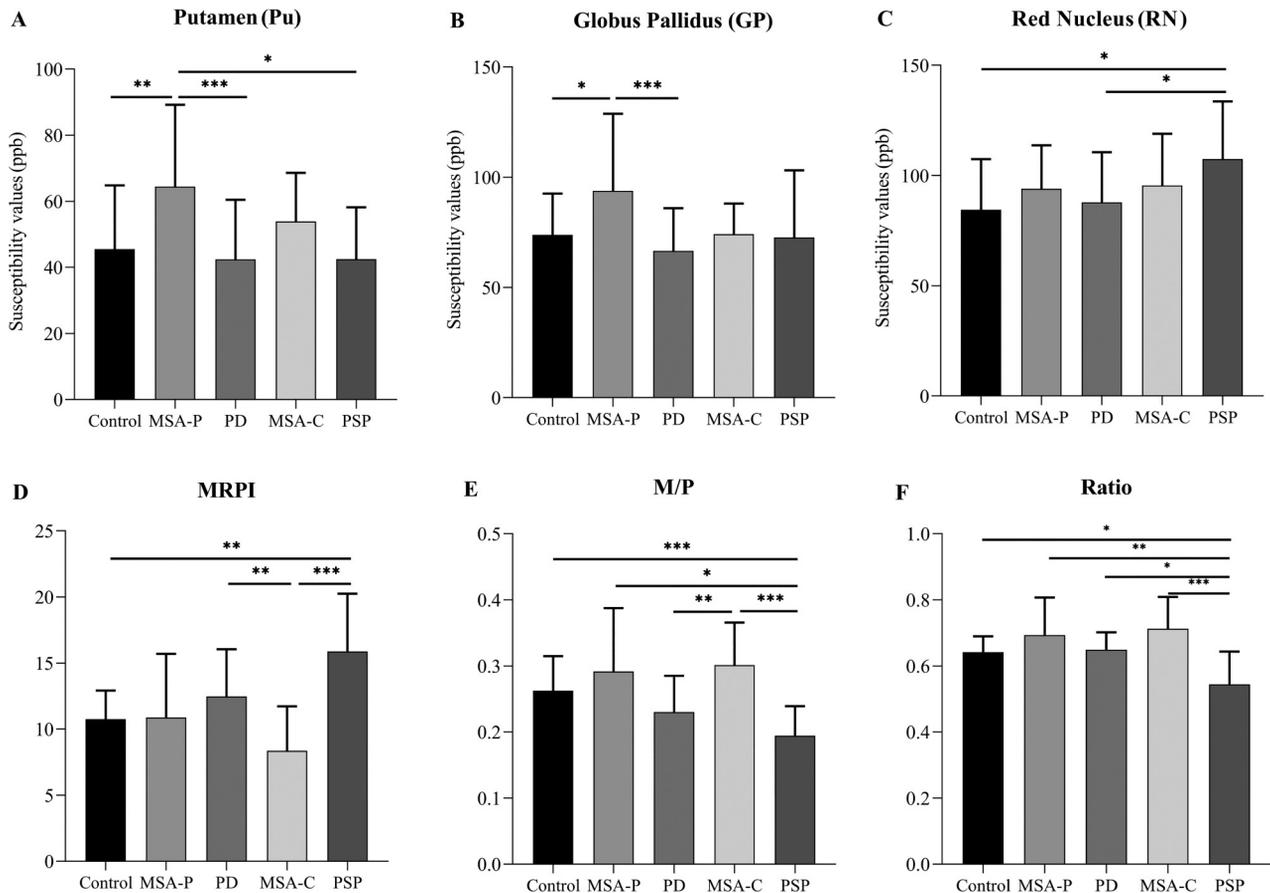
Age, disease duration, UPDRS-III score, and H-Y stage are quantitative data that are shown as mean ± standard deviation. Sex is qualitative data.

<sup>a</sup>1-way ANOVA was used for multiple comparisons of variance chi-square, 1-way ANOVA followed by post hoc Bonferroni test.

<sup>b</sup>Welch ANOVA was used if multiple comparisons were not variance-robust, Welch ANOVA followed by post hoc Tamhane's test.

<sup>c</sup>Significantly highest.

Sex by using chi-square test statistical methods.



**FIG 3.** Intergroup comparison of QSM susceptibility values (A, B, C) and morphometric parameters (D, E, F). Only significantly different ROIs and morphometrics are represented. Bars represent mean ± standard error. \*\*\* $P \leq .001$ ; \*\* $0.001 < P \leq .01$ ; \* $0.01 < P \leq .05$ . One-way ANOVA or Welch test was used for the comparison of multiple copies, and Bonferroni or Tamhane's test correction was applied for post hoc 2-by-2 comparisons. ppb = parts per billion.

of the PD group was higher than the MSA-C group ( $P = .004$ ) (Fig 3D). However, The M/P of the PD was lower than MSA-C ( $P = .01$ ). The M/P of the PSP group was lower than the Control ( $P = .001$ ), MSA-P ( $P = .01$ ), and MSA-C ( $P < .001$ ) groups (Fig 3E). The Ratio of the PSP group was lower than the Control ( $P = .03$ ), PD ( $P = .01$ ), MSA-P ( $P = .004$ ), and MSA-C ( $P = .001$ ) groups (Fig 3F).

### The Best Indicators to Identify Parkinsonism

Supplemental Data demonstrates the diagnostic value of variables with significant differences when used alone or in combination to

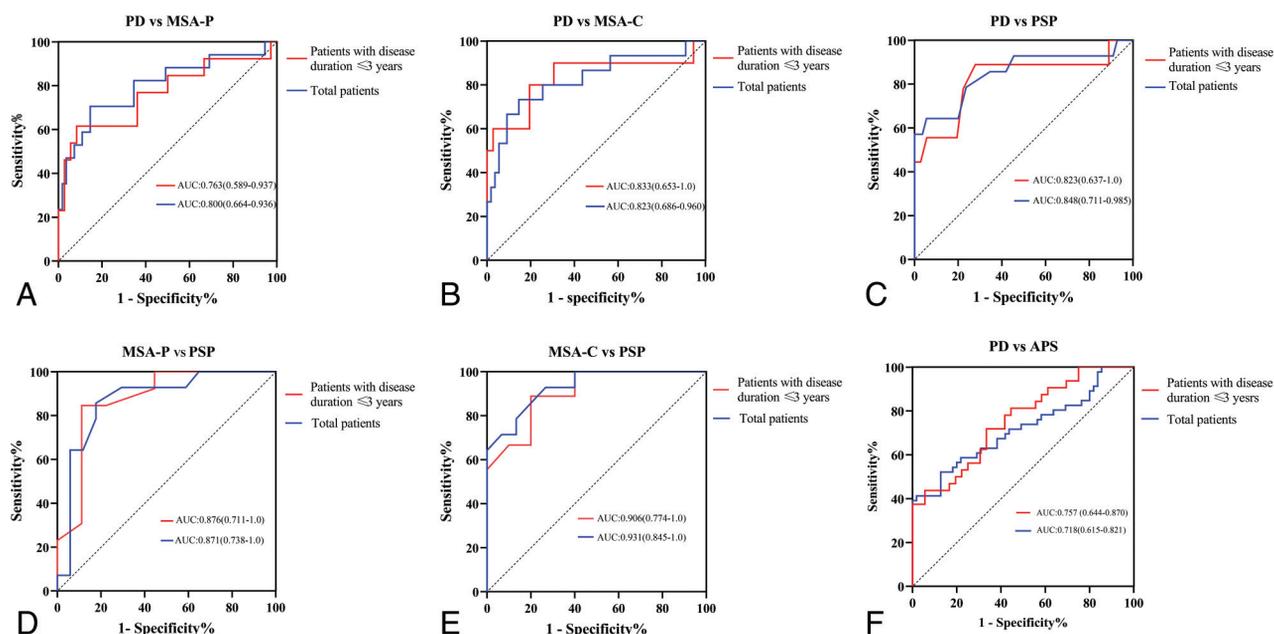
differentiate between 2 diseases. A combination indicator is a new test variable created by including variables that are significantly different between groups in a binary logistic regression. The choice of the number of independent variables and covariates to be included was based on the sample size, and the best indicators for distinguishing between PD and APS and diseases in APS were selected on a combined merit basis.

The cutoff values were determined using Youden's Index, which maximizes the sum of sensitivity and specificity for optimal diagnostic performance. The AUC of Pu combined with GP to differentiate PD from MSA-P was 0.800 (95% CI:

**Table 2: Receiver operator characteristics in the differential diagnosis of parkinsonism<sup>a</sup>**

	Auc	95% CI	P value	Youden Index	Cutoff value	Sensitivity	Specificity
PD vs MSA-P							
QSM (Pu, GP)	0.800	0.664–0.936	<.001	0.561	0.28	70.6%	85.5%
PD vs MSA-C							
Morphometry (MRPI, M/P)	0.823	0.686–0.960	<.001	0.588	0.26	73.3%	85.5%
PD vs PSP							
Ratio	0.848	0.711–0.985	<.001	0.588	0.57	94.5%	64.3%
MSA-P vs PSP							
Ratio	0.871	0.738–1.0	<.001	0.753	0.64	82.4%	92.9%
MSA-C vs PSP							
M/P	0.931	0.845–1.0	<.001	0.662	0.26	73.3%	92.9%
PD vs APS							
QSM (Pu, GP, RN) + M/P	0.718	0.615–0.821	<.001	0.395	0.65	41.3%	98.2%

<sup>a</sup> Youden Index = Sensitivity + Specificity - 1; Cutoff value = Max (Youden Index) corresponding to the test variable.



**FIG 4.** Differential diagnosis of ROC of PD, MSA-P, MSA-C, and PSP in total patients versus patients with disease duration of 3 years or less.

0.664–0.936). The AUC for MRPI combined with M/P to distinguish PD from MSA-C was 0.823 (95% CI: 0.686–0.960). The optimal indicators used in PD to identify different subtypes of MSA vary, but their AUC and specificity consistently exceed 80%. The ratio alone differentiated PSP from PD or MSA-P from PSP, and the AUCs were above 0.800. The AUC for M/P alone to distinguish MSA-C from PSP was 0.931 (95% CI: 0.845–1.0). The AUC was 0.718 (95% CI: 0.615–0.821) when combining the QSM (Pu, GP, RN) with the morphometric parameter (M/P) to differentiate PD from APS (Table 2).

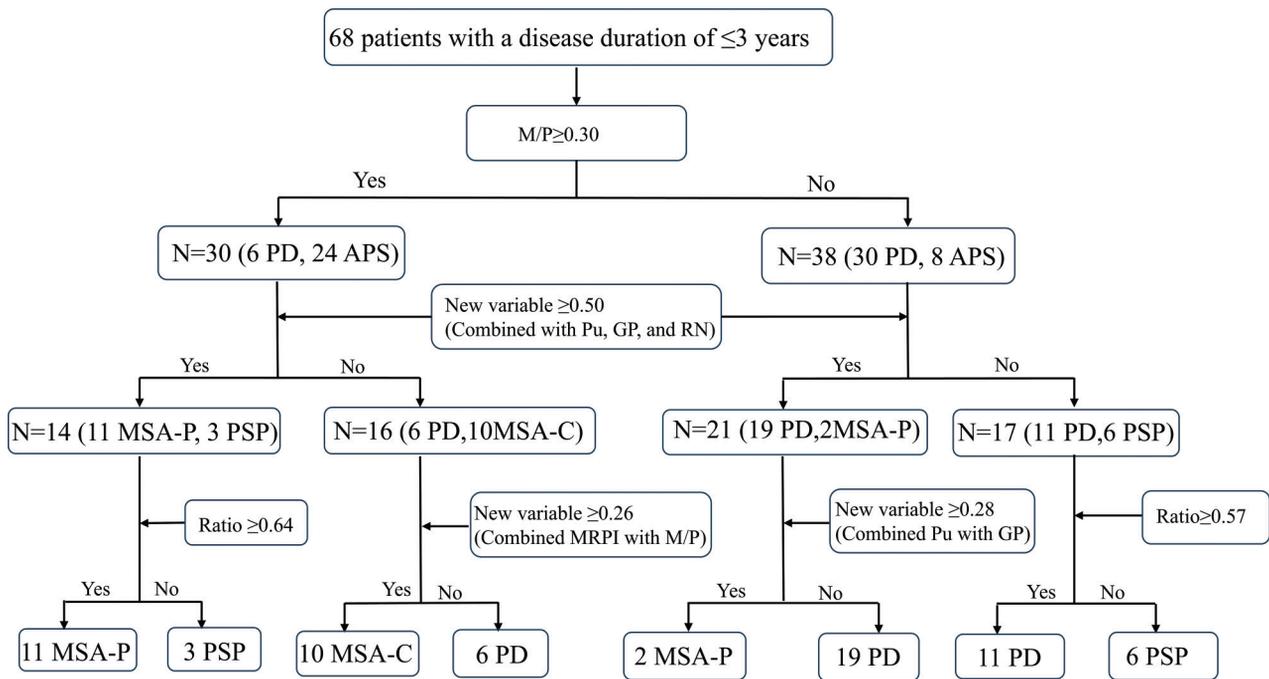
#### The ROC Curve of Patients with Disease Duration ≤ 3 Years

In addition to the full cohort analysis, a subgroup analysis was performed for patients with ≤ 3 years of disease duration. The model based on the best indicator already identified in Table 2 was validated in a subgroup of patients with a disease duration of ≤ 3 years. A total of 68 patients were seen for further analyses, including 36 with PD, 13 with MSA-P, 10 with MSA-C, and 9 with PSP.

Compared with all patients, the AUC improved in differentiating PD from MSA-C or APS patients with a disease duration of ≤ 3 years (Fig 4B and 4F). However, the AUC decreased in differentiating PD from MSA-P or PSP patients with a disease course of ≤ 3 years (Fig 4A and 4C). Additionally, the AUC improved in distinguishing PSP from MSA-P patients with a disease duration of ≤ 3 years, while it decreased in distinguishing PSP from MSA-C (Fig 4D and 4E). This subgroup exhibited comparable diagnostic performance across all methods, with AUC values closely aligned with those of the full cohort (Fig 4). These findings demonstrate the robustness of the diagnostic approaches in early disease stages.

#### Diagnostic Algorithm

A diagnostic algorithm representing the diagnostic classification of disease duration ≤ 3 years according to the best indicators cutoffs is shown in Table 3 (Fig 5). First, the combination of the M/P cutoff value of 0.30 and the cutoff value of 0.50 for QSM (Pu, GP, RN) susceptibility values was applied to distinguish PD and APS



**FIG 5.** Diagnostic algorithm representing the diagnostic classification of disease duration  $\leq 3$  years according to the best indicator cutoffs in Table 3. The new variable is formed by the combination of several variables in parentheses.

patients with a disease duration of  $\leq 3$  years. Second, the differentiation is based on the cutoff value of the best indicator for differential diagnosis between different diseases. Finally, we found that the diagnostic indicators proposed in this study could distinguish these patients with different diseases of  $\leq 3$  years duration.

### Statistical Measurements

The ICCs for intrarater and interrater reliability of the measurement of the susceptibility values of ROIs and morphometric parameters varied from 0.980 to 0.997 and 0.956 to 0.979, respectively (Supplemental Data).

## DISCUSSION

In this single-center retrospective study, we found that QSM has significant diagnostic value in differentiating between PD and MSA-P, and morphometry shows good diagnostic value in distinguishing between PD, MSA-C, and PSP. QSM combined with morphometry had the best diagnostic value to differentiate MSA-P from PSP and to differentiate PD from APS. Further subgroup analyses showed that the best indicators had equally high differential diagnostic values in patients with a disease duration of 3 years or less. Specifically, we emphasize its focus on the diagnostic value of QSM and morphometry in a northeastern Chinese population, as well as the inclusion of a healthy control group, which strengthens the validity of our findings.

Previous studies have shown that it is possible to differentiate between PD, APS, ET, and healthy controls using DTI, neuronal synapse orientation dispersion, attenuation imaging, and free water imaging.<sup>30,31</sup> It has also been shown that automated analysis of mean diffusivity in the middle cerebellar peduncle and shell nuclei has very high diagnostic accuracy in distinguishing between PD and MSA.<sup>32</sup> However, most of these are single MRI

studies, and the present study used a combination of 2 quantitative MRIs to further improve the differential diagnosis between PD and APS and diseases within APS.

In our study, the method of ROI confirmation in QSM not only considered the midbrain portion but also included the basal ganglia region and part of the cerebellar nuclei, given the wider lesions in MSA and PSP. For morphologic measurements, we not only included MRPI and M/P, which are 2 commonly used morphologic measurements but also newly included Ratio, which is a new indicator, and demonstrated that Ratio has a good diagnostic value in the differentiation of PD from PSP and MSA-P from PSP. This is novel compared with previous studies.

### MRI Findings

Previous studies have found iron deposition in the substantia nigra of PD patients,<sup>33</sup> and this finding was also affirmed in the present study. But the susceptibility values of substantia nigra were not statistically significant among all groups, which suggests that iron deposition in the substantia nigra does not provide diagnostic value in differentiating PD from APS. The comparison of multiple groups revealed that the MSA-P group had significantly higher susceptibility values in the Pu and the GP than the control and PD groups in this study. The susceptibility values in the RN of the PSP group were higher than the control and PD groups. Ahmed et al<sup>34</sup> showed that the increased susceptibility values in the Pu and GP of MSA patients might be due to the aggregation of  $\alpha$ -synuclein in glial cells in those regions. Yamamoto et al<sup>35</sup> suggested that iron stored in ferritin may promote the aggregation of  $\tau$  protein in PSP patients. In addition, we found that iron deposition in the basal ganglia was the least in PD patients and the most in MSA patients. According to the functional neuroanatomy model of the basal ganglia, the activity of the GABAergic

pathway from the Pu to the GP is increased in PD patients.<sup>36</sup> Experimental evidence that the utilization of GABA leads to a reduction in the iron concentration<sup>37</sup> suggests that the lower iron concentration in the Pu and GP of the PD brain may be attributable to an alteration in the GABA metabolism.<sup>38</sup>

A recent clinical drug trial study of deferiprone in PD patients<sup>39</sup> showed that deferiprone significantly reduced substantia nigra iron levels. In addition, an *in vitro* animal study<sup>40</sup> showed that iron chelators protect the neurons *in vitro* by inhibiting ferroptosis pathways. Our study showed that iron deposition in MSA and PSP is more widely distributed, reflecting differences in the anatomic patterns of iron deposition associated with neurodegenerative processes. Thus, studies on inhibiting disease progression by chelating iron are not only applicable to PD patients but may also be beneficial to MSA and PSP patients.

We also found that the mean MRPI of PSP patients is the highest, while the M/P and Ratio are the lowest among the groups. Conversely, the MSA-C patients exhibited the lowest mean MRPI while having the highest M/P and Ratio. The differences in morphometry may relate to the pathologic changes of diseases.<sup>2,4</sup> Furthermore, most previous studies<sup>41,42</sup> have concentrated on 2 morphometric parameters: MRPI and M/P. However, our study demonstrates that the morphometric parameter Ratio is highly effective in distinguishing PSP from PD. Our findings are consistent with the pathologic findings of midbrain atrophy, which are often associated with  $\tau$  protein deposition in patients with advanced-stage PSP.<sup>43</sup>

To validate our results further, we also performed an automated analysis of QSM that showed similar findings (Supplemental Data), which further demonstrated the reliability of our manual measurements versus the automated processing. However, we apologize that we were not able to automate the processing of morphometry, which is still being explored further.

### **Diagnostic Value of the Best Indicators for Identifying Parkinsonism**

We simplified the model by reducing the number of independent variables and covariates used in the subgroup analyses to address the potential overfitting of PSP with MSA-P or MSA-C groups. This approach ensures a more reliable interpretation of the results without overfitting due to small sample sizes. The sample sizes of the remaining subgroups were largely in line with conventional requirements concerning the number of variables they included.

The identified cutoff values have potential clinical relevance, as they provide practical thresholds for differentiating between disease groups. For instance, a cutoff of 0.28 allows for improved early detection of disease MSA-P, which is critical for timely intervention. The AUC of Pu and GP combined to distinguish PD and MSA-P can reach 0.800 in this study. Marxreiter et al<sup>44</sup> study was consistent with our findings. However, Sjöström et al<sup>45</sup> discovered that the susceptibility value of GP for identifying PD and MSA was only moderate or insignificant, likely due to differences in calculation methods and software used for plotting ROIs.

This study also found that combining MRPI with M/P can effectively distinguish between PD and MSA-C, and the AUC can reach 0.823. This finding is supported by the study of Möller et al.<sup>41</sup>

However, morphometry failed to distinguish between PD and MSA-P. This limitation arises because the typical morphologic changes in MSA-C involve atrophy of the cerebellum and pons, whereas most patients with MSA-P primarily show posterior putamen changes.<sup>46,47</sup>

When RN was utilized to differentiate PD from PSP, the AUC was 0.71, with a sensitivity of 64.3% and a specificity of 74.5%, in agreement with Sjöström et al.<sup>48</sup> Ratio alone was used to effectively differentiate PD from PSP (AUC = 0.848,  $P < .001$ ). In addition, we confirmed that morphometric parameters reflecting atrophy of the midbrain are the best indicators for distinguishing PSP from other groups.<sup>27,41</sup>

In the current study, we found that the specificity of morphometric parameters in distinguishing PSP from MSA-P or PSP from MSA-C was 92.9%. Previous study<sup>49</sup> affirmed the high diagnostic value of M/P in distinguishing PD from APS patients, but it was not in our study ( $P > .05$ ). The sample sizes may account for these contradictory results. The specificity of combining QSM (Pu, GP, RN) with morphometric parameters (M/P) to differentiate PD from APS reached 98.2%.

Finally, based on the best indicators, we propose a diagnostic algorithm validated in patients with disease duration  $\leq 3$  years (Fig 5). According to our results, the combination of QSM and morphometry can improve the differential diagnosis among PD, MSA-P, MSA-C, and PSP. The subgroup analysis for patients with  $\leq 3$  years of disease duration highlight the clinical utility of the proposed diagnostic methods in early-stage disease. Consistent diagnostic performance in this subgroup underscores the potential for timely and accurate disease differentiation, which is critical for guiding early therapeutic interventions. Therefore, combining QSM with morphometry can provide a basis for treatment by identifying the extent of abnormal iron deposition and brain atrophy in PD with APS disease at an early stage.

**Limitations.** This study also has several limitations. First, manual measurements are inherently subject to human error and variability, which can lead to inconsistencies, especially when different operators are involved. However, we minimized measurement errors by fixing the number of layers and averaging multiple measurements. Second, manual measurements are time-consuming, but they allow for more precise identification of anatomic structures. To further validate our results, we also performed an automated analysis of QSM, which showed similar findings (Supplemental Data), demonstrating the reliability of both manual and automated methods. Third, the number of patients with MSA and PSP is limited, which is mainly related to the low prevalence of these diseases. More samples will be accumulated, and multicenter studies will be conducted to validate the diagnostic value further in future studies. In addition, patients with CBS and nondegenerative Parkinsonism were not included in this study population.

### **CONCLUSIONS**

Our study showed that QSM and morphometry have strengths, respectively, in the differential diagnosis of PD from APS and diseases within APS. This study provides more imaging diagnostic evidence for the differentiation between PD, MSA-P, MSA-C, and PSP.

## ACKNOWLEDGMENTS

The work was supported by the First Hospital of Jilin University. We thank Professor Yanhua Wu for her help in statistical analysis and Wei Sun for her help in automated image analysis.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Postuma RB, Berg D, Stern M, et al. **MDS clinical diagnostic criteria for Parkinson's disease.** *Mov Disord* 2015;30:1591–601 CrossRef Medline
2. Wenning GK, Stankovic I, Vignatelli L, et al. **The Movement Disorder Society criteria for the diagnosis of multiple system atrophy.** *Mov Disord* 2022;37:1131–48 CrossRef Medline
3. Höglinger GU, Respondek G, Stamelou M; Movement Disorder Society-endorsed PSP Study Group, et al. **Clinical diagnosis of progressive supranuclear palsy: the movement disorder society criteria.** *Mov Disord* 2017;32:853–64 CrossRef Medline
4. Dickson DW. **Parkinson's disease and parkinsonism: neuropathology.** *Cold Spring Harb Perspect Med* 2012;2:a009258 CrossRef Medline
5. Massey LA, Micallef C, Paviour DC, et al. **Conventional magnetic resonance imaging in confirmed progressive supranuclear palsy and multiple system atrophy.** *Mov Disord* 2012;27:1754–62 CrossRef Medline
6. Meijer FJA, van Rumund A, Fasen BACM, et al. **Susceptibility-weighted imaging improves the diagnostic accuracy of 3T brain MRI in the work-up of parkinsonism.** *AJNR Am J Neuroradiol* 2015;36:454–60 CrossRef Medline
7. Halliday GM, Holton JL, Revesz T, et al. **Neuropathology underlying clinical variability in patients with synucleinopathies.** *Acta Neuropathol* 2011;122:187–204 CrossRef Medline
8. Paviour DC, Price SL, Jahanshahi M, et al. **Longitudinal MRI in progressive supranuclear palsy and multiple system atrophy: rates and regions of atrophy.** *Brain* 2006;129:1040–49 CrossRef Medline
9. Nicoletti G, Fera F, Condino F, et al. **MR imaging of middle cerebellar peduncle width: differentiation of multiple system atrophy from Parkinson disease.** *Radiol* 2006;239:825–30 CrossRef Medline
10. Williams DR, Holton JL, Strand C, et al. **Pathological tau burden and distribution distinguishes progressive supranuclear palsy-parkinsonism from Richardson's syndrome.** *Brain* 2007;130:1566–76 CrossRef Medline
11. Paviour DC, Price SL, Stevens JM, et al. **Quantitative MRI measurement of superior cerebellar peduncle in progressive supranuclear palsy.** *Neurology* 2005;64:675–79 CrossRef Medline
12. Tsuboi Y, Slowinski J, Josephs KA, et al. **Atrophy of superior cerebellar peduncle in progressive supranuclear palsy.** *Neurology* 2003;60:1766–69 CrossRef Medline
13. Fearnley JM, Lees AJ. **Ageing and Parkinson's disease: substantia nigra regional selectivity.** *Brain* 1991;114(5):2283–301 CrossRef Medline
14. Ide S, Murakami Y, Futatsuya K, et al. **Usefulness of olfactory bulb measurement in 3D-FIESTA in differentiating Parkinson disease from atypical parkinsonism.** *AJNR Am J Neuroradiol* 2024;45:1141–52 CrossRef Medline
15. Wang Z-L, Yuan L, Li W, et al. **Ferroptosis in Parkinson's disease: glia-neuron crosstalk.** *Trends Mol Med* 2022;28:258–69 CrossRef Medline
16. Biondetti E, Santin MD, Valabrègue R, et al. **The spatiotemporal changes in dopamine, neuromelanin and iron characterizing Parkinson's disease.** *Brain* 2021;144:3114–25 CrossRef Medline
17. Depierreux F, Parmentier E, Mackels L, et al. **Parkinson's disease multimodal imaging: f-DOPA PET, neuromelanin-sensitive and quantitative iron-sensitive MRI.** *NPJ Parkinsons Dis* 2021;7:57 CrossRef Medline
18. Boelmans K, Holst B, Hackius M, et al. **Brain iron deposition fingerprints in Parkinson's disease and progressive supranuclear palsy.** *Mov Disord* 2012;27:421–27 CrossRef Medline
19. Wang Y, Butros SR, Shuai X, et al. **Different iron-deposition patterns of multiple system atrophy with predominant parkinsonism and idiopathic Parkinson diseases demonstrated by phase-corrected susceptibility-weighted imaging.** *AJNR Am J Neuroradiol* 2012;33:266–73 CrossRef Medline
20. Frosini D, Ceravolo R, Tosetti M, et al. **Nigral involvement in atypical parkinsonisms: evidence from a pilot study with ultra-high field MRI.** *J Neural Transm (Vienna)* 2016;123:509–13 CrossRef Medline
21. Meijer FJ, Steens SC, van Rumund A, et al. **Nigrosome-1 on susceptibility weighted imaging to differentiate Parkinson's Disease from atypical parkinsonism: an in vivo and ex vivo pilot study.** *Pol J Radiol* 2016;81:363–69 CrossRef Medline
22. Langkammer C, Schweser F, Krebs N, et al. **Quantitative susceptibility mapping (QSM) as a means to measure brain iron? A post mortem validation study.** *NeuroImage* 2012;62:1593–99 CrossRef Medline
23. Goetz CG, Fahn S, Martinez-Martin P, et al. **Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (MDS-UPDRS): process, format, and clinimetric testing plan.** *Mov Disord* 2007;22:41–47 CrossRef Medline
24. Hoehn MM, Yahr MD. **Parkinsonism: onset, progression and mortality.** *Neurology* 1967;17:427–42 CrossRef Medline
25. Cosottini M, Frosini D, Pesaresi I, et al. **MR imaging of the substantia nigra at 7 T enables diagnosis of Parkinson disease.** *Radiol* 2014; 271:831–38 CrossRef Medline
26. Mazzucchi S, Frosini D, Costagli M, et al. **Quantitative susceptibility mapping in atypical Parkinsonisms.** *Neuroimage Clin* 2019;24: 101999 CrossRef Medline
27. Quattrone A, Nicoletti G, Messina D, et al. **MR imaging index for differentiation of progressive supranuclear palsy from Parkinson disease and the Parkinson variant of multiple system atrophy.** *Radiol* 2008;246:214–21 CrossRef Medline
28. Massey LA, Jäger HR, Paviour DC, et al. **The midbrain to pons ratio: a simple and specific MRI sign of progressive supranuclear palsy.** *Neurology* 2013;80:1856–61 CrossRef Medline
29. Parkinson's Disease and Movement Disorders Group. **Neurology Division, Chinese Medical Association, China, and Specialized Committee on Parkinson's Disease and Movement Disorders, Neurologists Branch, Chinese Physicians Association (CPA). Chinese clinical diagnostic criteria for progressive supranuclear palsy.** *Chinese J Neurol* 2016;49:272–76.
30. Prodoehl J, Li H, Planetta PJ, et al. **Diffusion tensor imaging of Parkinson's disease, atypical parkinsonism, and essential tremor.** *Mov Disord* 2013;28:1816–22 CrossRef Medline
31. Mitchell T, Archer DB, Chu WT, et al. **Neurite orientation dispersion and density imaging (NODDI) and free-water imaging in Parkinsonism.** *Hum Brain Mapp* 2019;40:5094–107 CrossRef Medline
32. Krismer F, Beliveau V, Seppi K, et al. **Automated analysis of diffusion-weighted magnetic resonance imaging for the differential diagnosis of multiple system atrophy from Parkinson's disease.** *Mov Disord* 2021;36:241–45 CrossRef Medline
33. Reimão S, Ferreira S, Nunes RG, et al. **Magnetic resonance correlation of iron content with neuromelanin in the substantia nigra of early-stage Parkinson's disease.** *Eur J Neurol* 2016;23:368–74 CrossRef Medline
34. Ahmed Z, Asi YT, Sailer A, et al. **The neuropathology, pathophysiology and genetics of multiple system atrophy.** *Neuropathol Appl Neurobiol* 2012;38:4–24 CrossRef Medline
35. Yamamoto A, Shin R-W, Hasegawa K, et al. **Iron (III) induces aggregation of hyperphosphorylated tau and its reduction to iron (II) reverses the aggregation: implications in the formation of neurofibrillary tangles of Alzheimer's disease.** *J Neurochem* 2002;82:1137–47 CrossRef Medline

36. Hill JM. **Iron concentration reduced in ventral pallidum, globus pallidus, and substantia nigra by GABA-transaminase inhibitor, gamma-vinyl GABA.** *Brain Res* 1985;342:18–25 CrossRef Medline
37. Christine CW, Aminoff MJ. **Clinical differentiation of parkinsonian syndromes: prognostic and therapeutic relevance.** *Am J Med* 2004; 117:412–19 CrossRef Medline
38. Graham JM, Paley MN, Grünewald RA, et al. **Brain iron deposition in Parkinson's disease imaged using the PRIME magnetic resonance sequence.** *Brain* 2000;123(12):2423–31 CrossRef Medline
39. Devos D, Labreuche J, Rascol O; FAIRPARK-II Study Group, et al. **Trial of deferiprone in Parkinson's Disease.** *N Engl J Med* 2022; 387:2045–55 CrossRef Medline
40. Zeng X, An H, Yu F, et al. **Benefits of iron chelators in the treatment of Parkinson's disease.** *Neurochem Res* 2021;46:1239–51 CrossRef Medline
41. Möller L, Kassubek J, Südmeyer M, et al. **Manual MRI morphometry in Parkinsonian syndromes.** *Mov Disord* 2017;32:778–82 CrossRef Medline
42. Nigro S, Morelli M, Arabia G, et al. **Magnetic resonance parkinsonism index and midbrain to pons ratio: which index better distinguishes progressive supranuclear palsy patients with a low degree of diagnostic certainty from patients with Parkinson disease?** *Parkinsonism Relat Disord* 2017;41:31–36 CrossRef Medline
43. Williams DR, Lees AJ. **Progressive supranuclear palsy: clinicopathological concepts and diagnostic challenges.** *Lancet Neurol* 2009; 8:270–79 CrossRef Medline
44. Marxreiter F, Lambrecht V, Mennecke A, et al. **Parkinson's disease or multiple system atrophy: potential separation by quantitative susceptibility mapping.** *Ther Adv Neurol Disord* 2023;16: 17562864221143834 CrossRef Medline
45. Mazzucchi S, Frosini D, Costagli M, et al. **Quantitative susceptibility mapping in atypical Parkinsonisms.** *Neuroimage Clin* 2019;24: 101999 CrossRef Medline
46. Chougar L, Pyatigorskaya N, Lehericy S. **Update on neuroimaging for categorization of Parkinson's disease and atypical parkinsonism.** *Curr Opin Neurol* 2021;34:514–24 CrossRef Medline
47. Pellecchia MT, Barone P, Mollica C, et al. **Diffusion-weighted imaging in multiple system atrophy: a comparison between clinical subtypes.** *Mov Disord* 2009;24:689–96 CrossRef Medline
48. Sjöström H, Granberg T, Westman E, et al. **Quantitative susceptibility mapping differentiates between parkinsonian disorders.** *Parkinsonism Relat Disord* 2017;44:51–57 CrossRef Medline
49. Oba H, Yagishita A, Terada H, et al. **New and reliable MRI diagnosis for progressive supranuclear palsy.** *Neurology* 2005;64:2050–55 CrossRef Medline

# Antiamyloid Therapy and Cerebral Blood Flow Changes on MRI: A Potential Longitudinal Biomarker of Treatment Response?

Andres Ricaurte-Fajardo,  Jana Ivanidze,  Deborah Zhang, Meem Mahmud, Weiye Yasen,  Lisa Ravdin,  Silky Pahlajani,  Mony de Leon, Anna S. Nordvig, and  Gloria C. Chiang



## ABSTRACT

**SUMMARY:** Amyloid-targeting therapy has recently become widely available in the United States for the treatment of patients with symptomatic mild Alzheimer disease (AD). At present, there are no biomarkers that have been clinically validated to assess treatment response in routine clinical practice; longitudinal amyloid PET could play a role but is not cost-effective. This report presents a case series of 6 patients with AD, whose amyloid positivity was confirmed by PET or CSF biomarkers, who underwent baseline and longitudinal arterial spin-labeling MR imaging (ASL-MR) as part of Food and Drug Administration–mandated, clinical standard-of-care, noncontrast MR monitoring to assess for amyloid-related imaging abnormalities (ARIA). We and others have previously reported that ASL-MR can screen for neurodegenerative disease as a proxy for FDG-PET and can be easily added on as a cost-effective, repeatable method to monitor post therapy changes. This series highlights varied CBF changes in response to lecanemab therapy. For instance, Cases 1, 3, and 5 showed increased CBF after multiple infusions, with subjective cognitive improvement in Case 1 and improved MoCA scores in Case 3. Case 2 showed improved CBF initially before the fifth infusion, but this returned to baseline in the subsequent study, with no cognitive improvement over the course of therapy. Cases 4 and 6 have demonstrated no substantial changes in regional CBF thus far on therapy, with cognitive decline in Case 4. This case series underscores the potential utility of ASL-MR as an adjunct sequence to current imaging protocols to monitor treatment response to antiamyloid therapy.

**ABBREVIATIONS:**  $A\beta$  = amyloid- $\beta$ ; AD = Alzheimer disease; ARIA = amyloid-related imaging abnormalities; ASL-MR = arterial spin-labeling MR imaging; MoCA = Montreal Cognitive Assessment

For the past year, lecanemab has been increasingly available in the United States for the treatment of symptomatic mild cognitive impairment or mild dementia due to Alzheimer disease (AD),<sup>1</sup> and donanemab was recently approved for the same indication.<sup>2</sup> To qualify for therapy, patients have to undergo amyloid PET or CSF assays<sup>1</sup> to confirm the presence of amyloid pathology. Lecanemab is a recombinant humanized IgG1 monoclonal antibody that targets amyloid oligomers, protofibrils, and insoluble fibrils.<sup>3</sup> It binds preferentially to protofibrils, which are a high

molecular weight form of soluble amyloid and are a component of amyloid plaques.<sup>3</sup> Despite its availability, at present, there are no biomarkers that have been clinically validated to assess treatment response in routine clinical practice; longitudinal amyloid PET could play a role but is not cost-effective. While lecanemab has shown a modest impact on cognition over 18 months, slowing cognitive decline by 27%, there is minimal detectable cognitive impact in the first 6 months of treatment, and very little is known about the brain's response to therapy. Therefore, tracking the effects of early medication remains elusive for clinicians.<sup>4</sup>

The US Food and Drug Administration (FDA) mandates serial noncontrast MRI scans to monitor for amyloid-related imaging abnormalities (ARIA) while receiving bimonthly lecanemab treatment (before the fifth, seventh, and 14th infusions).<sup>1</sup> Additionally, an MRI scan at week 52 (before the 26th infusion) is suggested, particularly for *apolipoprotein E*  $\epsilon 4$  carriers, who have an increased risk of ARIA, and for those who had ARIA in prior studies.<sup>1</sup>

Arterial spin-labeling MRI (ASL-MR) uses magnetically labeled water (ie, blood) as an endogenous tracer to measure CBF in the brain.<sup>5-7</sup> Unlike other perfusion imaging techniques that use <sup>15</sup>O-PET or dynamic imaging with gadolinium contrast, ASL-MR does not require an external tracer or ionizing radiation.<sup>7</sup> CBF, estimated by ASL-MR, is closely linked to brain

Received November 6, 2024; accepted after revision December 30.

From the Department of Radiology (A.R.-F., J.I., S.P., M.d.L., G.C.C.), Brain Health Imaging Institute, and Department of Neurology (D.Z., M.M., W.Y., L.R., A.S.N.), Weill Cornell Medicine, New York Presbyterian Hospital, New York, New York; and Department of Neurology (A.R.-F.), Pontificia Universidad Javeriana, Bogota, Colombia.

Anna S. Nordvig and Gloria C. Chiang contributed equally to this article.

This research was supported in part by the following grants: National Institutes of Health/National Institute on Aging R01AG068398 (G.C.C.), R01AG080011 (G.C.C.), R01 AG085972 (G.C.), R56 AG058913 (M.d.L.), and RFI AG057570 (M.d.L.), Sonia Jaye and Edward Barsukov Foundation (A.S.N.), and Feil II Clinical Scholar Award (A.S.N.).

Please address correspondence to Gloria C. Chiang, MD, Department of Radiology, Division of Neuroradiology, Weill Cornell Medicine, NY Presbyterian Hospital, 525 East 68th St, Starr Pavilion, Box 141, New York, NY 10065; e-mail: gcc9004@med.cornell.edu; @GloriaChiangMD; @WCMRadiology

 Indicates article with supplemental data.

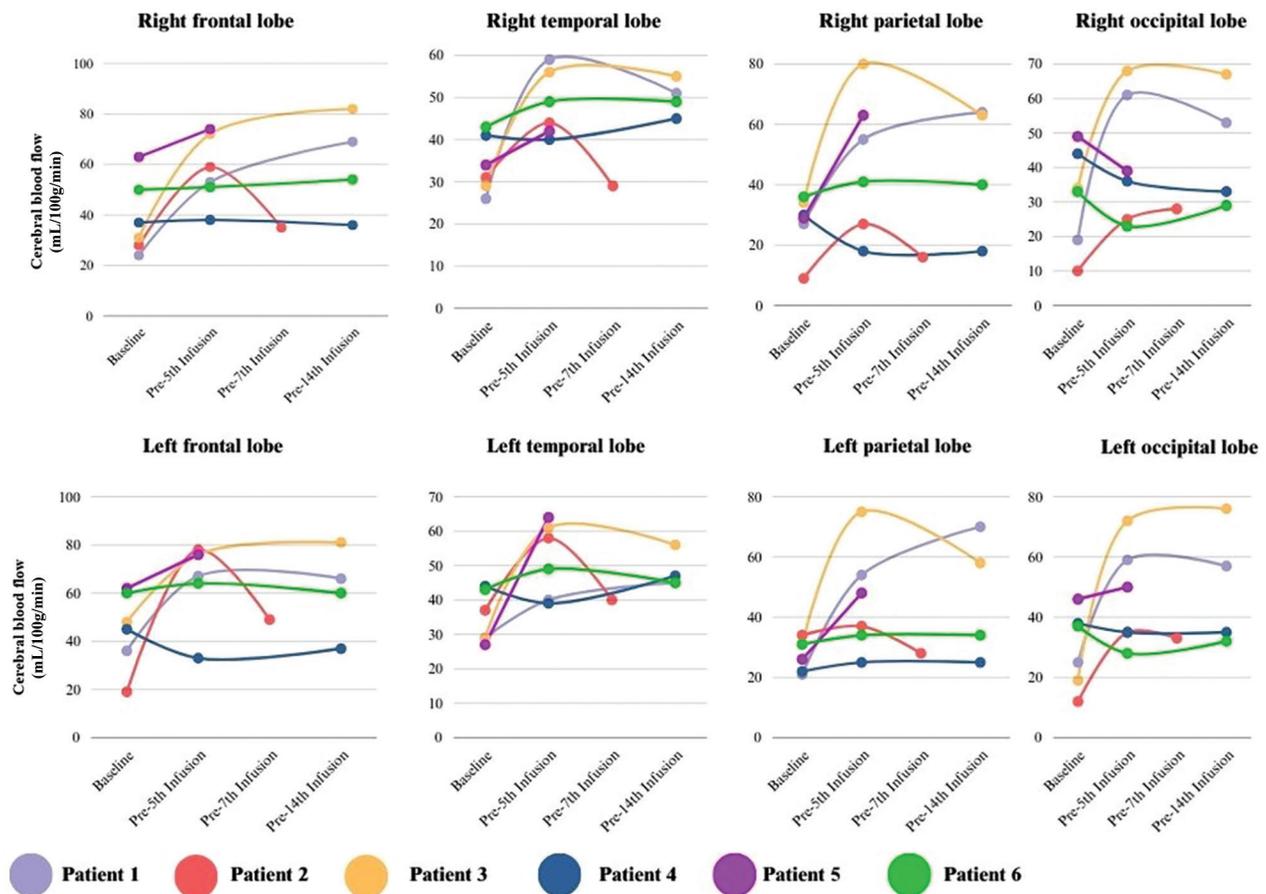
<http://dx.doi.org/10.3174/ajnr.A8654>

**Patient characteristics, including history of vascular risk factors**

Age	Confirmation of Alzheimer Pathology	APOE Genotype	Baseline MoCA score	Hypertension	Dyslipidemia	Diabetes
77	Amyloid PET	E3/E3	23	Yes	Yes	No
59	CSF biomarkers	E3/E3	20	No	No	No
74	Amyloid PET	E3/E3	22	No	No	No
62	Amyloid PET	E3/E3	16 (MMSE 22)	No	Yes	No
79	Amyloid PET	E3/E3	21	No	Yes	No
68	CSF biomarkers	E4/E4	14 (CDR 1)	Yes	Yes	No

Note:—MMSE indicates Mini-Mental State Examination.

**Cerebral Blood Flow Trends Across Infusions**

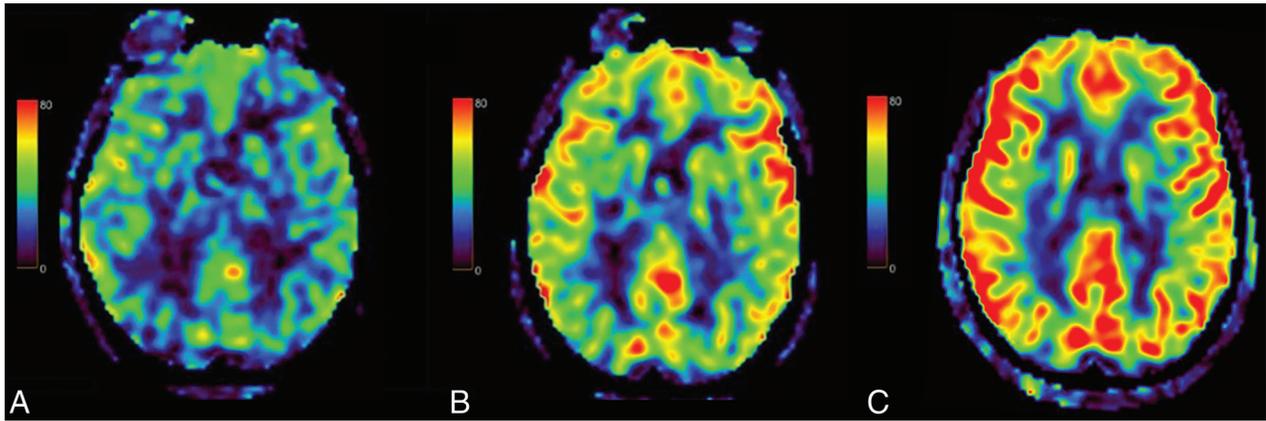


**FIG 1.** Line graphs showing trends in CBF values across time points. Data are shown for the 6 patients included in this case series, with measurements in the frontal, temporal, parietal, and occipital lobes, right or left.

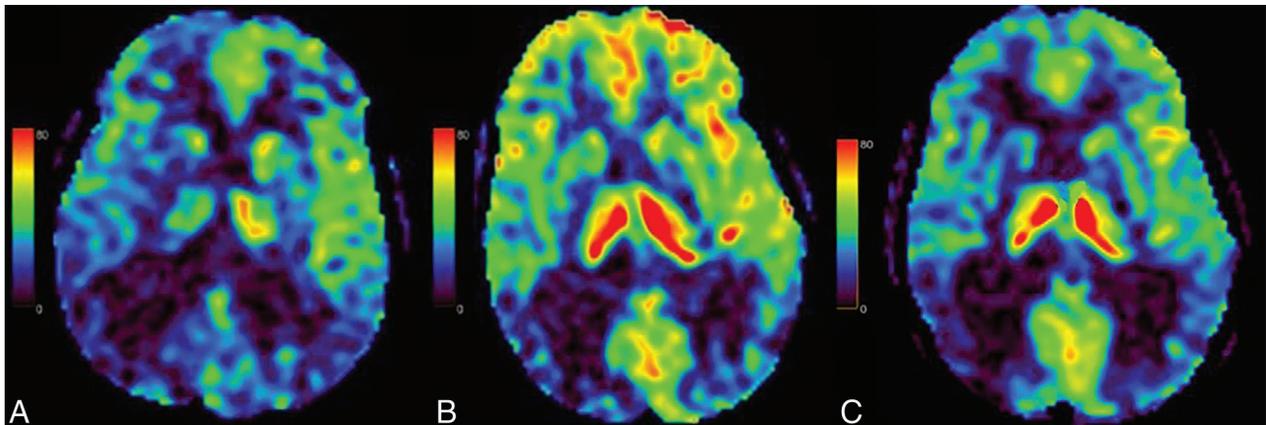
metabolism and associates with FDG-PET, thus representing an efficient, low-cost, noninvasive, and quantitative method for evaluating brain function.<sup>5-7</sup>

In this report, we present 6 cases of patients with AD, whose amyloid positivity was confirmed by PET or CSF biomarkers, who underwent serial ASL-MR while on lecanemab therapy (Table). All patients were imaged on the same clinical 3T Signa Architect MRI scanner (GE Healthcare). The 3D pseudo continuous ASL-MR clinical product sequence had the following parameters: post-labeling delay 2025 ms, TR 4876, TE 53.6, 38 × 4 mm axial slices, FOV 24, matrix size 512; the scanning time was 4 minutes and 24 seconds (Supplemental Data). CBF maps were processed by using the AW Server Version 3.2 extension 4.0 (GE Healthcare), set to a range of 0–80 mL/100 g/min by using

the rainbow color scheme. Images were visually interpreted in native space. Regions-of-interest were also placed on the CBF maps to assess quantitative changes over the course of therapy (Fig 1, Supplemental Data). All patients had well-controlled, stable comorbidities, and all followed treatment guidelines for initiation of therapy and monitoring. All cases were followed clinically up to the time of the writing of this manuscript, which was at least to the 14th infusion for Cases 1, 3, and 4. We hypothesized that anti-amyloid therapy would lead to improved cerebral perfusion, as seen on ASL-MR, potentially due to clearance of both parenchymal and vascular amyloid.<sup>8</sup> Here, we present the CBF maps of these patients at baseline and while receiving lecanemab, illustrating the potential utility of ASL-MR in monitoring the effects of lecanemab on brain perfusion.



**FIG 2.** Patient 1 CBF maps. A, Baseline CBF map, before lecanemab treatment, demonstrating hypoperfusion, most prominently in the left frontal and bilateral parietal lobes. Posttreatment CBF map, obtained before the administration of the fifth dose of lecanemab (B), reveals a marked overall increase in cerebral perfusion in all brain regions, which increases further on the CBF map before the 14th dose (C).



**FIG 3.** Patient 2 CBF maps. A, Baseline CBF map showing hypoperfusion in the bilateral frontal, parietal, and occipital lobes, more pronounced on the right. B, CBF map before the fifth dose of lecanemab, demonstrating marked improvement in cerebral perfusion, especially in the bilateral frontal lobes. C, CBF map before the seventh dose of lecanemab, showing decreased overall perfusion compared with the prior MRI and no substantial change from baseline.

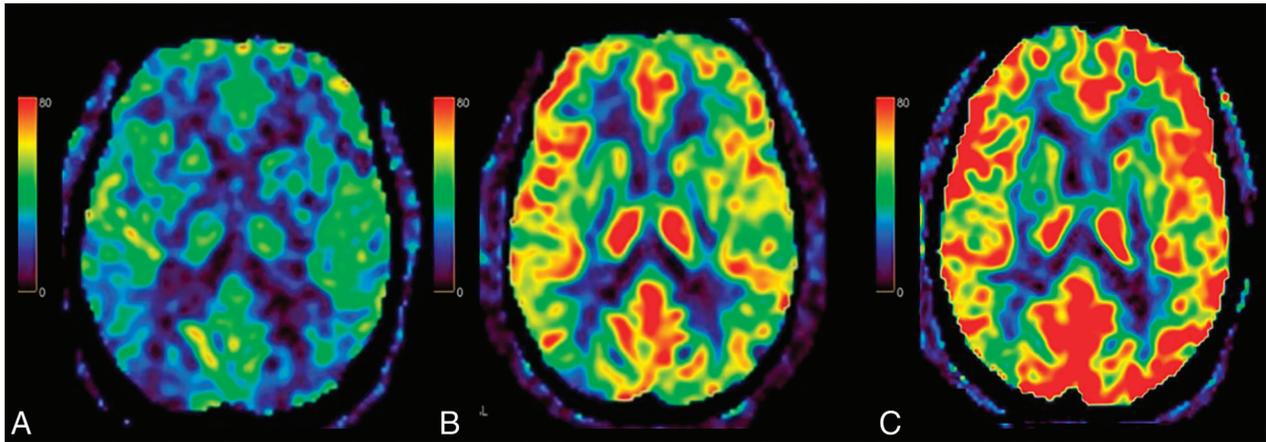
### CASE 1

A 77-year-old right-handed woman with a medical history of hyperlipidemia, hypertension, and heart failure, with an *APOE*  $\epsilon 3/\epsilon 3$  genotype, presented with a 1.5-year history of progressive short-term memory loss, word-finding difficulty, occasional disorientation, and computer apraxia. Symptoms possibly began after parotid gland surgery, during which she experienced mild postoperative delirium. Her baseline Montreal Cognitive Assessment (MoCA) score was 23 out of 30. FDG-PET showed symmetric temporoparietal hypometabolism, a typical pattern for AD (Supplemental Data), and ASL-MR revealed decreased CBF in the left parietal and temporal lobes, as well as the bilateral frontal lobes (Fig 2). A positive amyloid PET confirmed the diagnosis of AD (Supplemental Data). She started treatment with lecanemab. ASL-MR performed at the time of the monitoring MRI scans, before the fifth, seventh, and 14th infusions, showed no ARIA and markedly improved CBF compared with baseline, most notably before the 14th infusion (Fig 2). The patient and family reported subjective cognitive improvement, although the MoCA score remained unchanged at 23/30 after 6 months of therapy. Notably, there was no change in mild

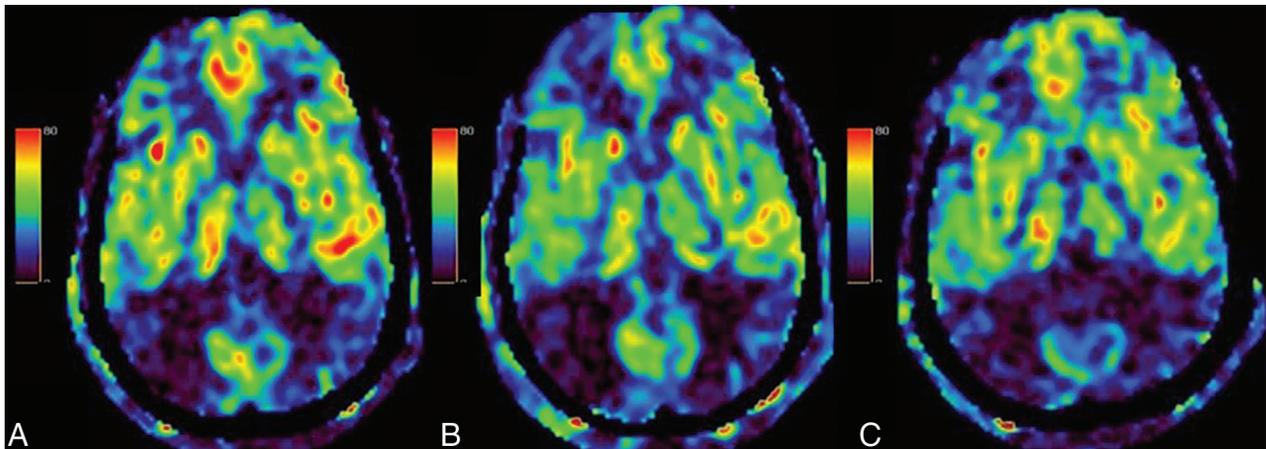
volume loss or minimal burden of white matter hyperintensities over the course of therapy (Supplemental Data).

### CASE 2

A 59-year-old right-handed woman presented with an 8-month history of progressive short-term memory loss, confusion, and increased anxiety. She struggled with tasks such as understanding the calendar and organizing simple work assignments. The patient had normal walking, with no shuffling or dragging of feet, and reported no urinary incontinence. Her initial MoCA score was 20/30; her *APOE* genotype was  $\epsilon 3/\epsilon 3$ . FDG-PET showed temporoparietal hypometabolism, and MRI showed marked volume loss and decreased CBF in the bilateral temporal and parietal lobes, as well as the frontal lobes (Fig 3, Supplemental Data). CSF analysis confirmed the diagnosis of early-onset AD. She began treatment with donepezil, followed by lecanemab. The patient reported feeling less anxious and more socially active, although cognitive challenges persisted. The CBF map showed marked improvement on the scan before the fifth infusion, but the CBF then decreased on the subsequent



**FIG 4.** Patient 3 CBF maps. A, Baseline CBF map showing reduced cerebral blood flow in the bilateral parietal lobes, as well as the frontal lobes, to a lesser extent. B, CBF map before the fifth dose of lecanemab, showing improvement in blood flow in all brain regions. C, CBF map before the 14th dose of lecanemab, showing further increased blood flow, compared with prior monitoring and baseline CBF maps.



**FIG 5.** Patient 4 CBF maps. A, Baseline CBF map showing hypoperfusion in the bilateral parietal lobes, as well as the right frontal lobe, to a lesser extent. B, CBF map before the fifth dose of lecanemab, with no substantial change in overall cerebral perfusion compared with baseline. C, CBF map before the 14th dose of lecanemab again showing no notable change in cerebral perfusion compared with the previous studies.

scan before the seventh dose (Fig 3). There was no evidence of ARIA; the minimal burden of white matter hyperintensities stayed stable (Supplemental Data).

### CASE 3

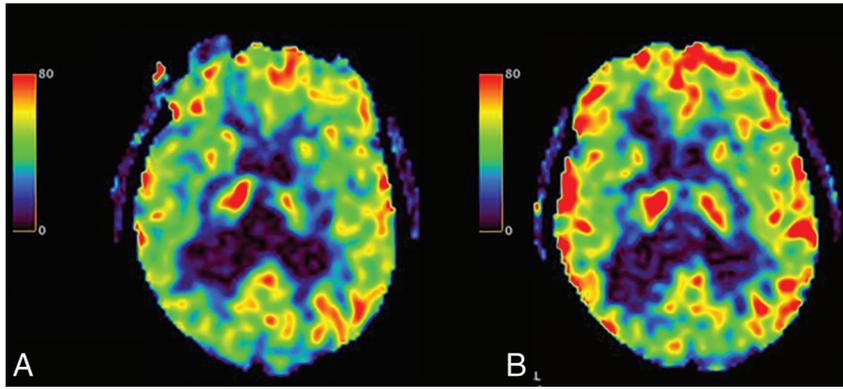
A 74-year-old right-handed man presented with a 10-year history of slow progressive memory decline, life-long depression, monoclonal gammopathy of unknown significance with paresthesias, prostate cancer, and sleep apnea. Three paternal relatives were reported to have a history of cognitive impairment. His baseline MoCA score was 22/30 with deficits in executive function, visuospatial tasks, and short-term recall; his *APOE* genotype was  $\epsilon 3/\epsilon 3$ . Baseline FDG-PET showed frontal hypometabolism initially favored to represent mild frontotemporal dementia, while MRI showed mild global volume loss and decreased CBF in the bilateral parietal, temporal, and frontal lobes (Fig 4, Supplemental Data). CSF biomarkers were indeterminate, but subsequent amyloid and  $\tau$  PET scans were positive, confirming AD (Supplemental Data). Monitoring MRI scans before the fifth, seventh, and 14th lecanemab infusions showed no

evidence of ARIA and continued improvement in CBF (Fig 4) despite no change in mild volume loss and minimal burden of white matter hyperintensities (Supplemental Data). The patient reported improved working memory and overall cognitive function; the prelecanemab MoCA score on donepezil (26/30) improved after 6 months of lecanemab therapy to 29/30.

### CASE 4

A 62-year-old right-handed man presented with progressive cognitive decline over 2.5 years and worsened anxiety after a suspected COVID infection. His medical history included hypertension and concussion. Cognitive assessments revealed deficits in visuospatial organization, working memory, attention, and orientation. Neuropsychological testing indicated performance below expectation, compared with education-based norms, in multiple cognitive domains, with substantial deficits in visuo-perceptual ability, working memory, and verbal fluency. His *APOE* genotype was  $\epsilon 3/\epsilon 3$ .

FDG-PET showed temporoparietal hypometabolism, and MRI showed decreased CBF bilaterally in the temporal, parietal, and frontal lobes (Fig 5). CSF biomarkers and amyloid PET confirmed



**FIG 6.** Patient 5 CBF maps. A, Baseline CBF map showing cerebral blood flow to be within normal limits. B, CBF map before the fifth dose of lecanemab showing increased cerebral perfusion.

AD with an associated inflammatory profile, including elevated oligoclonal bands, possibly triggered by COVID infection. Donepezil, memantine, and desvenlafaxine somewhat improved symptoms. Five days of high-dose methylprednisolone very briefly improved executive function, mood, and alertness. He had 1 microhemorrhage at baseline, but he did not develop ARIA. ASL-MR before the fifth, seventh, and 14th dose monitoring scans for lecanemab did not show improvement in CBF compared with baseline (Fig 5). Clinical history indicated a slight decrease in cognitive function; the prelecanemab MMSE of 22/30 decreased slightly to 21/30 after 9 months of lecanemab therapy. There was no change in mild-moderate volume loss (Supplemental Data). He had no white matter hyperintensities at baseline or on subsequent MRI scans.

### CASE 5

A 79-year-old right-handed woman presented with short-term memory loss for 2 years, with mild disorientation and irritability over the past 3 years. Her medical history included Paget disease and hyperlipidemia. Her mother had “AD-like” dementia in her eighties. Cognitive assessments revealed difficulties with short-term memory, working memory, visuospatial tasks, phonemic fluency, and naming. Her initial MoCA score was 21/30. ASL-MR and FDG-PET suggested AD, particularly with decreased CBF and metabolism in the temporal lobes (Fig 6, Supplemental Data). Amyloid PET confirmed the AD diagnosis (Supplemental Data). Monitoring MRI scans before the fifth and seventh lecanemab infusions showed improvement in CBF and no ARIA (Fig 6). Cognition was reported as stable by the family. Marked volume loss and mild white matter hyperintensity burden remained stable (Supplemental Data).

### CASE 6

A 68-year-old man presented with progressive forgetfulness and personality changes over 5 years, including increased jealousy and irritability. His medical history included hypertension, hyperlipidemia, and type 2 diabetes. Cognitive assessment showed a MoCA score of 14/30, affected by poor vision and low educational attainment; his Clinical Dementia Rating (CDR) should be

capitalized (it’s a specific validated tool) was 1.0. Family history included a mother with early-onset cognitive impairment and a brother with late-life paranoia.

ASL-MR and FDG-PET were concordant in suggesting AD (Supplemental Data); CSF biomarkers confirmed the diagnosis. The patient received initial treatment with donepezil and olanzapine with some improvement. Lecanemab therapy was then initiated. Monitoring MRI scans before the fifth and seventh infusions showed no ARIA and unchanged CBF (Supplemental Data); the patient’s family felt his cognition had remained stable to

slightly improved. Mild volume loss and minimal white matter hyperintensity burden remain stable (Supplemental Data).

### DISCUSSION

The accumulation of amyloid- $\beta$  ( $A\beta$ ) extracellularly in brain parenchyma and along the walls of cerebral blood vessels is a well-described pathologic characteristic of AD.<sup>9</sup>  $A\beta$  deposition in the walls and smooth muscle cells of cerebral arteries and arterioles, predominantly  $A\beta$ -40, the short form of  $A\beta$ , leads to decreased CBF.<sup>9,10</sup> The cases presented in this series highlight how CBF changes on ASL-MR could reflect clearance of  $A\beta$  on lecanemab therapy, thereby raising the possibility that ASL-MR could provide an early predictive biomarker of treatment response.

Similar to FDG-PET, longitudinal decline in CBF is associated with cognitive decline,<sup>11-13</sup> likely because of the tight coupling of blood flow, synaptic activity, and glucose metabolism.<sup>14</sup> This series demonstrates that ASL-MR can provide valuable insights into the neurovascular effects of lecanemab within 2–3 months of therapy, with 3 patterns of CBF change demonstrated: continued improvement (Cases 1, 3, 5), immediate increase that reverts to baseline (Case 2), and no improvement (Cases 4 and 6). ASL-MR can thus potentially serve as a noninvasive, cost-effective method for tracking treatment progress before the slowing of cognitive decline can be detected.<sup>9</sup>

In cases of increased CBF, antibody clearance by lecanemab may effectively clear vascular amyloid.<sup>8,15</sup> Preclinical investigations have provided evidence that monoclonal antibody therapies can protect neurons from  $A\beta$ -induced apoptosis, leading to neuronal viability and improved brain perfusion.<sup>16</sup> CBF increases may also reflect increased synaptic activity after amyloid clearance and/or vascular changes, such as clearance of fibrinogen/amyloid clotting complexes,<sup>17</sup> further leading to CSF clearance of toxic proteins and improved metabolic activity. Notably, the family of Case 1 reported cognitive improvement, and Case 3 demonstrated objective improvement in the MoCA.

In contrast, some patients did not exhibit substantial changes in CBF despite lecanemab therapy. Cases 4 and 6, for instance, demonstrated persistently low CBF on serial ASL-MR studies. In Case 2, we observed improved CBF before the fifth dose, but this did not persist on the subsequent ASL-MR. An initial increase in CBF may occur due to partial clearance of vascular amyloid.

However, in advanced disease, the extent of amyloid clearance may be insufficient to mitigate disease progression, as other non-amyloid mechanisms may also cause a decline in CBF. These findings highlight varied hemodynamic changes, possibly providing a metric of underlying treatment efficacy. Notably, Cases 2 and 4 had marked atrophy (Supplemental Data), which suggests that patients with more advanced disease may receive less benefit from therapy.<sup>1</sup>

Our series highlights the potential for CBF to serve as a tool for monitoring the neurovascular effects of lecanemab. Unlike structural MRI, which primarily is used to monitor for ARIA-E and ARIA-H, ASL-MR provides both a qualitative and quantitative assessment of cerebral perfusion, which can reflect functional brain changes in response to treatment. The noninvasive nature and relatively low cost of ASL-MR allows it to be easily added to the MRI studies that are already mandated by the FDA during monitoring, further enhancing its suitability for routine clinical use, particularly when repeated measures are necessary. Our series is unique in that serial ASL-MR scans were performed on the same exact MRI scanner, mitigating technical reasons for CBF changes. However, validation of these results in larger cohorts is warranted, and future research can evaluate whether these early changes in CBF can predict true treatment efficacy, as demonstrated by amyloid clearance and slowing of cognitive decline.

This study has several limitations. As a case series, these findings may not be generalizable to the broader population of patients with AD undergoing anti-amyloid therapies. Larger cohort studies are needed to validate these results and statistically evaluate the significance of CBF changes across serial ASL-MR. In addition, none of the patients presented had evidence of ARIA; future research will consider the potential effects of ARIA on CBF.

As noted in the Table, several of our patients had vascular comorbidities, which were stable but could potentially influence CBF. We chose not to exclude patients with these vascular comorbidities because patients with AD in real-world settings have high rates of these comorbidities and are being treated with anti-amyloid therapies. For example, 1 study reported that almost one-half of their treated patients with AD had hypertension, 24% had diabetes, and more than 70% had hypercholesterolemia.<sup>18</sup> By including such patients in our clinical case series, we show that CBF maps could be informative in these populations as well.

Finally, this study relied on manual ROI analysis, which can be subject to user variability, potential bias, and partial volume effects. However, we chose this form of analysis for easy translation into clinical practice.

Notably, we observed an improvement in cognitive symptoms in cases 1, 3, and 6, although anti-amyloid therapies are primarily designed to slow disease progression rather than directly improve clinical symptoms. Further longitudinal clinical follow-up could provide insights into these cognitive improvements. Future work could also determine whether CBF changes reflect changes in cerebral metabolism on FDG-PET or glymphatic clearance by using CSF dynamics or diffusion metrics.

## CONCLUSIONS

This case series highlights a potential role for noncontrast ASL-MR to detect changes in CBF in patients with AD

receiving lecanemab therapy. The varied response to therapy shown in this case series, although not yet well understood, suggests that CBF changes may provide an opportunity for personalized approaches to monitoring and treatment. ASL-MR offers a promising, noninvasive, and low-cost means to track these changes, which could enhance our understanding and management of AD in clinical practice. Further longitudinal studies are needed to determine whether early CBF improvements can predict better cognitive response to anti-amyloid and other upcoming forms of AD therapy.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Cummings J, Apostolova L, Rabinovici GD, et al. **Lecanemab: appropriate use recommendations.** *J Prev Alzheimers Dis* 2023;10:362–77 CrossRef Medline
2. FDA approves treatment for adults with Alzheimer's disease. US Food and Drug Administration. July 2, 2024; available from: <https://www.fda.gov/drugs/news-events-human-drugs/fda-approves-treatment-adults-alzheimers-disease>
3. Soderberg L, Johannesson M, Nygren P, et al. **Lecanemab, aducanumab, and gantenerumab - binding profiles to different forms of amyloid-beta might explain efficacy and side effects in clinical trials for Alzheimer's disease.** *Neurotherapeutics* 2023;20:195–206 CrossRef Medline
4. van Dyck CH, Swanson CJ, Aisen P, et al. **Lecanemab in early Alzheimer's disease.** *N Engl J Med* 2023;388:9–21 CrossRef Medline
5. Lee K, Mahmud M, Marx D, et al. **Clinical arterial spin-labeling MR imaging to screen for typical and atypical neurodegenerative disease in the new era of Alzheimer treatment.** *AJNR Am J Neuroradiol* 2024;45:632–36 CrossRef Medline
6. Wierenga CE, Hays CC, Zlatař ZZ. **Cerebral blood flow measured by arterial spin labeling MRI as a preclinical marker of Alzheimer's disease.** *J Alzheimers Dis* 2014;42 Suppl 4:S411–19 CrossRef Medline
7. Wolk DA, Detre JA. **Arterial spin labeling MRI: an emerging biomarker for Alzheimer's disease and other neurodegenerative conditions.** *Curr Opin Neurol* 2012;25:421–28 CrossRef Medline
8. Zago W, Schroeter S, Guido T, et al. **Vascular alterations in PDAPP mice after anti-A $\beta$  immunotherapy: implications for amyloid-related imaging abnormalities.** *Alzheimers Dement* 2013;9:S105–15 CrossRef Medline
9. Uekawa K, Hattori Y, Ahn SJ, et al. **Border-associated macrophages promote cerebral amyloid angiopathy and cognitive impairment through vascular oxidative stress.** *Mol Neurodegener* 2023;18:73 CrossRef Medline
10. Kakuda N, Miyasaka T, Iwasaki N, et al. **Distinct deposition of amyloid-beta species in brains with Alzheimer's disease pathology visualized with MALDI imaging mass spectrometry.** *Acta Neuropathol Commun* 2017;5:73 CrossRef Medline
11. Raz L, Knoefel J, Bhaskar K. **The neuropathology and cerebrovascular mechanisms of dementia.** *J Cereb Blood Flow Metab* 2016;36:172–86 CrossRef Medline
12. Hanaoka T, Kimura N, Aso Y, et al. **Relationship between white matter lesions and regional cerebral blood flow changes during longitudinal follow up in Alzheimer's disease.** *Geriatr Gerontol Int* 2016;16:836–42 CrossRef Medline
13. Weijs RWJ, Shkredova DA, Brekelmans ACM, et al. **Longitudinal changes in cerebral blood flow and their relation with cognitive decline in patients with dementia: current knowledge and future directions.** *Alzheimers Dement* 2023;19:532–48 CrossRef Medline
14. Chen Y, Wolk DA, Reddin JS, et al. **Voxel-level comparison of arterial spin-labeled perfusion MRI and FDG-PET in Alzheimer disease.** *Neurology* 2011;77:1977–85 CrossRef Medline

15. Morgan D. **Mechanisms of Abeta plaque clearance following passive A beta immunization.** *Neurodegener Dis* 2005;2:261–66 CrossRef Medline
16. Qiao Y, Chi Y, Zhang Q, et al. **Safety and efficacy of lecanemab for Alzheimer's disease: a systematic review and meta-analysis of randomized clinical trials.** *Front Aging Neurosci* 2023;15:1169499 CrossRef Medline
17. Chen ZL, Singh PK, Calvano M, et al. **A possible mechanism for the enhanced toxicity of beta-amyloid protofibrils in Alzheimer's disease.** *Proc Natl Acad Sci U S A* 2023;120:e2309389120 CrossRef Medline
18. Shields LBE, Hust H, Cooley SD, et al. **Initial experience with lecanemab and lessons learned in 71 patients in a regional medical center.** *J Prev Alzheimers Dis* 2024;11:1549–62 CrossRef Medline

# Deep Invasion Volume of the Primary Nasopharyngeal Carcinoma Predicts Treatment Outcome

Qi Yong H. Ai, Ho Sang Leung, Frankie K.F. Mo, Lun M. Wong, Linfang Lan, Edwin P. Hui, Brigette B.Y. Ma, and Ann D. King



## ABSTRACT

**BACKGROUND AND PURPOSE:** Quantification of deep invasion of the primary tumor is a predictor of outcome in oral cancer, but its predictive value in nasopharyngeal carcinoma (NPC) is unknown. This study aimed to investigate deep invasion of the primary NPC by using volumetric measurements on MRI for the prediction of outcome.

**MATERIALS AND METHODS:** Retrospective review was conducted of 822 MRIs from patients with newly diagnosed nonmetastatic NPC with volumetric analysis of the primary tumor to obtain total primary tumor volume (PTV), deep invasion volume (DIV), and ratio of deep to the total primary tumor volume (DIVr). Optimal predictors were identified by the multivariable Cox regression and c-index correlating with disease-free survival (DFS), distant metastases-free survival (DMFS), and overall survival (OS).

**RESULTS:** High DIVr, DIV, and PTV significantly correlated with poor DFS, DMFS, and OS (all  $P < .01$ ); DIVr being the optimal measurement (hazard ratio = 3.234 for DFS, 3.409 for DMFS, and 3.184 for OS). Compared with the eighth edition American Joint Committee on Cancer (AJCC) T-category, DIVr showed modest improvement in c-indexes for predicting DFS (0.602 versus 0.620,  $P = .03$ ) and DMFS (0.597 versus 0.626,  $P < .01$ ), but not OS ( $P = .15$ ). The use of a DIVr-based T-category had similar survival prognostication to the eighth edition AJCC T-category although there was improved prediction in DMFS.

**CONCLUSIONS:** DIVr is a better predictor of outcome in NPC than PTV or DIV, with slightly superior performance to the eighth edition AJCC T-category especially for DMFS.

**ABBREVIATIONS:** AJCC = American Joint Committee on Cancer; DFS = disease-free survival; DIV = volume of deep invasion; DIVr = ratio of deep to the total primary tumor volume; DMFS = distant metastases-free survival; DOI = depth of invasion; NPC = nasopharyngeal carcinoma; OS = overall survival; PTV = primary tumor volume; UICC = International Union Against Cancer

Depth of invasion (DOI), defined as the invasion below the level of epithelial basement membrane, is an important predictor of outcome in head and neck cancer, especially in carcinoma of the tongue in which a greater DOI is associated

with an increased risk of nodal metastases and decreased survival.<sup>1</sup> The unidimensional measurement of the DOI on pathologic examination influences the approach to definitive cancer management treatment,<sup>2,3</sup> and is incorporated into the eighth edition International Union Against Cancer/American Joint Committee on Cancer (UICC/AJCC) staging system.<sup>2-4</sup> Other pathologic measurements include tumor thickness (ie, maximal length of tumor from surface to its deep invading edge)<sup>5</sup> while imaging-based estimation of the pathologic DOI is increasingly utilized to guide preoperative planning for neck dissection.<sup>6</sup>

Despite the success of using DOI in oral cavity carcinoma, the concept has not been investigated in regard to outcome in nasopharyngeal carcinomas (NPCs). This is primarily because this cancer is treated by radiation therapy or chemoradiation so that pathologic evaluation of DOI is not available. It is also because accurate evaluation by using unidimensional measurement is hampered by the shape of the primary NPC.<sup>7</sup> Of note, the deeply invading component of tumor is often irregular in shape as it spreads in multiple directions beyond the confines of the nasopharyngeal walls into the adjacent soft tissues and bone. One way

Received August 1, 2024; accepted after revision December 16.

From the Department of Health Technology and Informatics (Q.Y.H.A., L.L.), The Hong Kong Polytechnic University, Kowloon, Hong Kong S.A.R., China; and Departments of Imaging and Interventional Radiology (Q.Y.H.A., H.S.L., L.M.W., A.D.K.) and Clinical Oncology, State Key Laboratory of Translational Oncology, Sir Y.K. Pao Centre for Cancer (F.K.F.M., E.P.H., B.B.Y.M.), The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong S.A.R., China.

Ethics numbers: CRE-2019.709.

Qi Yong H. Ai and Ho Sang Leung contributed equally to this work.

The research was conducted in part at CUHK DIIR MRI Facility, which is jointly funded by Kai Chong Tong, HKSAR Research Matching Grant Scheme and the Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong.

Please address correspondence to Ann D. King, Department of Imaging and Interventional Radiology, Faculty of Medicine, The Chinese University of Hong Kong, Prince of Wales Hospital, 30-32 Ngan Shing St, Shatin, New Territories, Hong Kong S.A.R., China; e-mail: king2015@cuhk.edu.hk

Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8678>

## SUMMARY

**PREVIOUS LITERATURE:** Quantification of deep invasion of primary oral cavity squamous cell carcinoma is a predictor of nodal metastasis and outcome and is incorporated in the AJCC T-category for staging. Although PTV in NPC correlates with relapse and survival, the predictive value of the deep invasion volume in this cancer is unknown.

**KEY FINDINGS:** The volume of deep invasion in primary NPC, expressed as a ratio of deep to the total primary tumor volume (DIVr), correlated with the presence of nodal metastases and was the best predictor of outcome, especially distant metastases-free survival. Using a DIVr-based T-category slightly improved outcome prediction when compared with the current AJCC T-category.

**KNOWLEDGE ADVANCEMENT:** DIVr reflects the propensity for deep invasion and performed better than total PTV for predicting NPC outcome and may be useful for guiding posttreatment image surveillance for distant metastases, but the impact on overall survival does not justify changing the AJCC T-categorization for NPC staging.

to overcome the inaccuracy of a unidimensional measurement is to obtain the radiologic volume of deep invasion (DIV), but until now there has been a reluctance to use volumes because this measurement is labor intensive. However, artificial intelligence-based segmentation of NPC is advancing rapidly<sup>8-12</sup> and developments in this area are likely to make automatic volume measurements a feasible option in the future.

In this study we evaluate the deep invasion of the primary tumor by using a volumetric measurement to correlate with the presence of nodal metastasis on the initial scan and predict survival in patients with NPC treated with curative intent. Furthermore, we postulate that tumors with a greater propensity to deep invasion are more aggressive and more likely to have nodal metastases at initial presentation and relapse at distant sites. Therefore, we also measure the radiologic DIV as a ratio of deep to the total primary tumor volume (DIVr). Finally, we determine if the addition of DIV could improve the staging system by using the eighth edition of the UICC/AJCC for the outcome of NPC.

## MATERIALS AND METHODS

### Patients

This retrospective study was approved by the local institutional review boards. The study was designed following the Strengthening of Reporting of Observational Studies in Epidemiology statement and the checklist is provided in the Supplemental Data. All study procedures complied with the tenets of the Declaration of Helsinki 2013 and its later amendments. The study included 844 consecutive patients with newly diagnosed, nondisseminated, pathologically proved NPC, who underwent a pretreatment staging MRI examination of the head and neck at Prince of Wales Hospital, Hong Kong S.A.R., between January 2005 and December 2016. Patients were treated with curative intent by using intensity modulated radiation therapy with or without chemotherapy. Patients were excluded because they were 1) lost to follow-up ( $n = 12$ ), 2) treatment was incomplete ( $n = 8$ ), and 3) MRI was degraded by artifact ( $n=2$ ), leaving 822 patients for analysis.

### Imaging Acquisition

MRI examinations of the head and neck were performed on a 1.5T or 3T whole-body system (Philips Medical Systems). The protocol consisted of a minimum of an axial fat-suppressed T2-

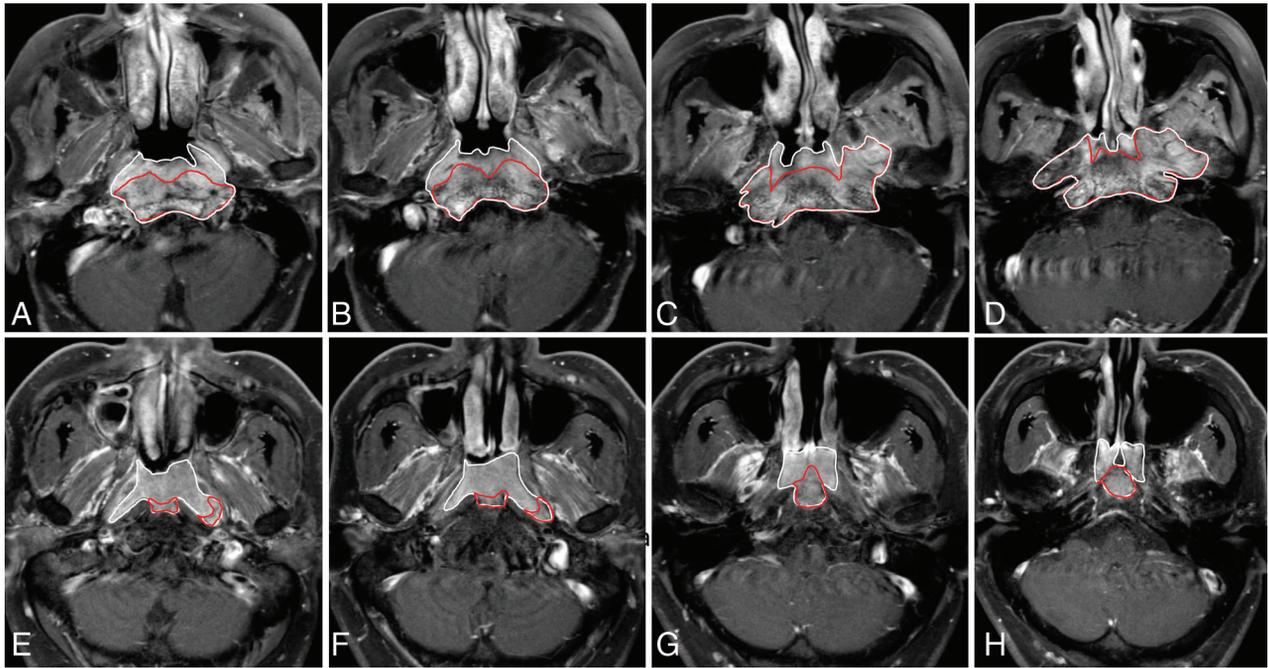
weighted sequence; an axial T1-weighted spin-echo sequence; axial and coronal T1-weighted spin-echo sequences with and without fat suppression, following a bolus injection of gadoterid acid (Dotarem; Guerbet) or gadolinium dimeglumine (Magnevist; Schering). Sequences were obtained by using a section thickness of 3–4 mm with no intersection gap, FOV of 23 cm, matrix varying from 256–800, and TR and TE dependent of field strength and sequence.

### MRI Analysis

Primary tumors were manually segmented on each of the slices of the axial postcontrast-enhanced T1-weighted images with the references to all available MRI sequences by using the open-source software ITK-SNAP (Version 3.4.0), by a researcher with 10 years' experience in head and neck MRI blinded to clinical outcome (Fig 1). The total primary tumor volume (PTV), DIV, and ratio of the DIV to the PTV (DIVr), expressed as a percentage ( $\text{DIVr} = \text{DIV}/\text{PTV} \times 100\%$ ), were obtained. The DIV was mapped separately to the PTV and the contouring of the DIV was based on the presence of invasion of structures deep to the nasopharynx beyond the fascia (ie, into tensor veli palatini muscle, parapharyngeal fat, medial and lateral pterygoid muscles, prevertebral muscles, skull base, paranasal sinuses, cervical spine, orbits, intracranial structures, and parotid gland). In addition, 100 patients were randomly selected and the manual segmentations of the PTV and DIV were obtained by a second observer, a radiologist with more than 3 years' experience in head and neck radiology for the evaluation of interobserver agreement for PTV measurements. The presence or absence of nodal metastases at initial diagnosis were also assessed based on recognized imaging criteria as follows: 1) retropharyngeal lymph nodes, cervical lymph nodes in the jugulodigastric region, and any other cervical lymph nodes with a minimal axial diameter of 5 mm, 11 mm, and 10 mm, respectively; 2) groups of 3 or more borderline lymph nodes with a minimal axial diameter of 8 mm; 3) lymph nodes of any size with extranodal extension or necrosis.<sup>13,14</sup>

### NPC Staging

Primary tumor (T) and metastatic nodes (N) were staged according to the eighth edition of the AJCC/UICC cancer staging manual.<sup>4,15</sup> The clinical overall stage was also obtained for analysis.



**FIG 1.** Pretreatment axial contrast-enhanced T1-weighted fat-suppressed MR images of 2 patients with T3 NPC. Both patients had similar total PTV (*white contours*) but different ratios of the DIVr, calculated by using the volume of deep invasion (DIV) (*red contours*)/PTV (*white contours*) \*100%. Primary tumors with a greater propensity for deep invasion had a poorer outcome as shown by comparing the patient with a tumor with a DIVr of 82.1% who developed distant metastases 0.8 years after treatment (A–D) to the patient with a tumor with a DIVr of 14.7% without disease recurrence 4.1 years after treatment (E–H).

### Treatment

The primary tumor and grossly enlarged lymph nodes received 66–70 Gy. Regions at risk of microscopic spread and the bilateral cervical lymphatics were selectively irradiated to 50–60 Gy. For chemotherapy, overall stage I and II disease without bulky cervical lymph nodes were treated by radiation therapy alone, while overall stages II with bulky cervical lymph nodes, III and IV diseases were treated by chemoradiotherapy. The standard chemotherapy regimen was concurrent cisplatin (mainly weekly low dose 40 mg/m<sup>2</sup>) with or without neoadjuvant or adjuvant chemotherapy.

### Follow-Up and Clinical Outcome

All patients underwent regular clinical follow-up after treatment once every 3 months for the first 12 months, then once every 6 months for the next 24 months, and then once yearly afterward until diagnosis of recurrence or death. Patients with suspected local or nodal recurrence or distant metastases underwent imaging by using conventional imaging techniques or FDG-PET/CT +/- biopsy, and clinical or radiologic follow-up. The survival end points in this study were disease-free survival (DFS), distant metastases-free survival (DMFS), and overall survival (OS), which were measured from the start of treatment to the date of any recurrence, last follow-up, or death. Recurrence at any site (local, nodal, or distant) was included in DFS, recurrence at distant site(s) was included in DMFS, and death from any cause was included in OS.

### Statistical Analysis

The PTV, DIV, and DIVr were correlated with the survival end points by using univariable Cox regression, followed by the multivariable Cox regression with the step-forward approach to select

the strongest predictor for assessment of the optimal measurement. Predictive values of the optimal volumetric measurement and that of the current AJCC T-category for DFS, DMFS, and OS were compared by using concordance statistics by the methods of both Harrell et al<sup>16</sup> and Uno et al.<sup>17</sup> The c-index is calculated as the ratio of the number of concordant pairs to the number of permissible pairs, and the index has showed a positive correlation with the capability in the discriminating between the paired groups, ie, a high c-index indicates a high level of discrimination. The method of Harrell et al<sup>16</sup> provides an overall measure of differences, and the method of Uno et al<sup>17</sup> estimates differences from baseline to a specific time point,<sup>18</sup> with 1000 bootstrapping to provide the biased-corrected c-index and corresponding 95% CIs.<sup>16,19</sup> The comparison in c-indexes between the optimal volumetric measurement and the eighth edition AJCC T-category for DFS, DMFS, and OS was performed according to the method by Uno et al.<sup>17</sup> A primary classification by using the optimal volumetric measurements of primary NPC for predicting risks of disease recurrence was proposed by using the classification and regression tree analysis. The survival rates of the proposed volumetric classification and the current AJCC T-category were evaluated by using the Kaplan-Meier analysis, and differences in survival rates were compared by log-rank test. Univariable and multivariable Cox regression analyses based on the eighth edition AJCC T-category and proposed classification by using new volume parameters, adjusting with nodal status, age, sex, and use of chemotherapy were performed. In addition, PTV, DIV, and DIVr were correlated with presence or absence of metastatic nodes (N0 versus n+) on the initial scan by using the univariable logistic regression. Interobserver agreement for the PTV measurements was evaluated

and the intraclass coefficients (ICCs) were calculated. A 2-tail *P* value of less than .05 was considered statistically significant. All analyses were performed by using SPSS (24.0 version, IBM) or SAS (9.4 version, SAS Institute).

**Table 1: Patient demographic characteristics and T- and N- categories and overall staging distribution**

	Number of Patients (Total n=822)
Age	
Median (range) (years)	53 (19–90)
Sex	
Man	631 (76.8%)
Woman	191 (23.2%)
Histologic types	
Squamous keratinized	3 (0.4%)
Squamous nonkeratinized	29 (3.5%)
Undifferentiated	790 (96.1%)
T-category (AJCC/UICC eighth edition)	
T1	312 (38.0%)
T2	110 (13.4%)
T3	250 (30.4%)
T4	150 (18.2%)
N-category (AJCC/UICC eighth edition)	
N0	231 (28.1%)
N1	321 (39.1%)
N2	185 (22.5%)
N3	85 (10.3%)
Overall stage (AJCC/UICC eighth edition)	
Stage I	105 (12.8%)
Stage II	195 (23.7%)
Stage III	309 (37.6%)
Stage IVA	213 (25.9%)
Volumetric measurements of the primary tumors (median and range)	
PTV (cm <sup>3</sup> )	13.3 (0.6–173.3)
DIV (cm <sup>3</sup> )	5.7 (0–169.0)
DIVr	45.5% (0%–98.2%)
Chemotherapy	
Yes	582 (70.8%)
No	240 (29.2%)
Outcome	
Death	268 (32.6%)
Died from NPC-related disease	176 (21.4%)
Died from non-NPC-related disease	92 (11.2%)
Disease recurrence	222 (27.0%)
Local recurrence	80 (9.7%)
Regional recurrence	45 (5.5%)
Distant metastases	146 (17.8%)
Follow-up	
Median follow-up (range) (year)	6.6 (0.3–15.1)

## RESULTS

### Patients

Patient demographics, histologic types of the tumor, T- and N-categories, and overall stages of treatment and posttreatment outcomes are shown in Table 1. The interobserver agreements were excellent for DIV (ICC of 0.86; 95% CI: 0.82–0.90) and PTV (ICC of 0.93; 95% CI: 0.90–0.95).

### PTV, DIV, and DIVr for Survival Prediction

The 5-year DFS, DMFS, and OS were 74.7%, 83.4%, and 78.1%, respectively. Table 2 shows univariable analysis of PTV, DIV, and DIVr, to correlate with DFS, DMFS, and OS. Results show that higher PTV, DIV, and DIVr correlated with poorer DFS, DMFS, and OS (all *P* < .01), with DIVr showing the strongest correlation (hazard ratio = 3.234 for DFS, 3.409 for DMFS, and 3.184 for OS) (Table 2).

### Predictive Value of DIVr and Comparison with the eighth edition AJCC T-Category

C-indexes of the AJCC T-category and DIVr for DFS, DMFS, and OS from baseline to 3-year, 5-year, and overall follow-up are shown in Table 3. Compared with the AJCC T-category, DIVr showed statistically higher overall c-indexes for DFS (0.602 versus 0.620, *P* = .03) DMFS (0.597 versus 0.626, *P* < .01), but not for OS (*P* = .15) (Table 3).

The DIVr thresholds identified for predicting survival were: DIVr-T1 = 0%; DIVr-T2: >0%, ≤60.0%; DIVr-T3: >60.0%, ≤80.0%; and DIVr-T4: >80.0%, resulting in 312, 169, 173, and 168 being reclassified to DIVr-T1, -T2, -T3, and -T4, respectively. For the AJCC-T2 category, 24 and 8 patients were reclassified by using DIVr to T3 and T4, respectively; for the AJCC-T3 category, 86 and 51 patients were reclassified by using DIVr to T2 and T4, respectively; for the AJCC-T4 category, 5 and 36 patients were reclassified by using DIVr to T2 and T3, respectively.

The survival rates of each risk group by using the proposed DIVr-based T-categories and the AJCC T-categories for DFS, DMFS, and OS are shown in Fig 2. Compared with the AJCC T-categories the proposed DIVr-based T-categories resulted in a slight improvement in survival prediction and the redistribution of numbers of patients in each of the categories (Fig 2). Similar to the AJCC T-category, the proposed DIVr-based T-category is an independent predictor of DFS, DMFS, and OS after controlling for N-category, age, sex, and use of chemotherapy (Supplemental Data). However, differences are observed in some subcategories, notably, DIVr-T4 shows significantly worse DMFS than the DIVr-T2–T3 (5-year DMFS for DIVr-T4 = 70.5%, log-rank test

**Table 2: Univariable analysis of 3 PTV measurements for predicting DFS, DMFS, and OS**

	DFS		DMFS		OS	
	HR (95% CI)	<i>P</i> Value	HR (95% CI)	<i>P</i> Value	HR (95% CI)	<i>P</i> Value
PTV	1.012 (1.008–1.016)	<.01	1.012 (1.007–1.017)	<.01	1.012 (1.008–1.016)	<.01
DIV	1.013 (1.009–1.018)	<.01	1.012 (1.007–1.017)	<.01	1.012 (1.008–1.016)	<.01
DIVr	3.234 (2.207–4.739)	<.01	3.409 (2.120–5.482)	<.01	3.184 (2.250–4.505)	<.01

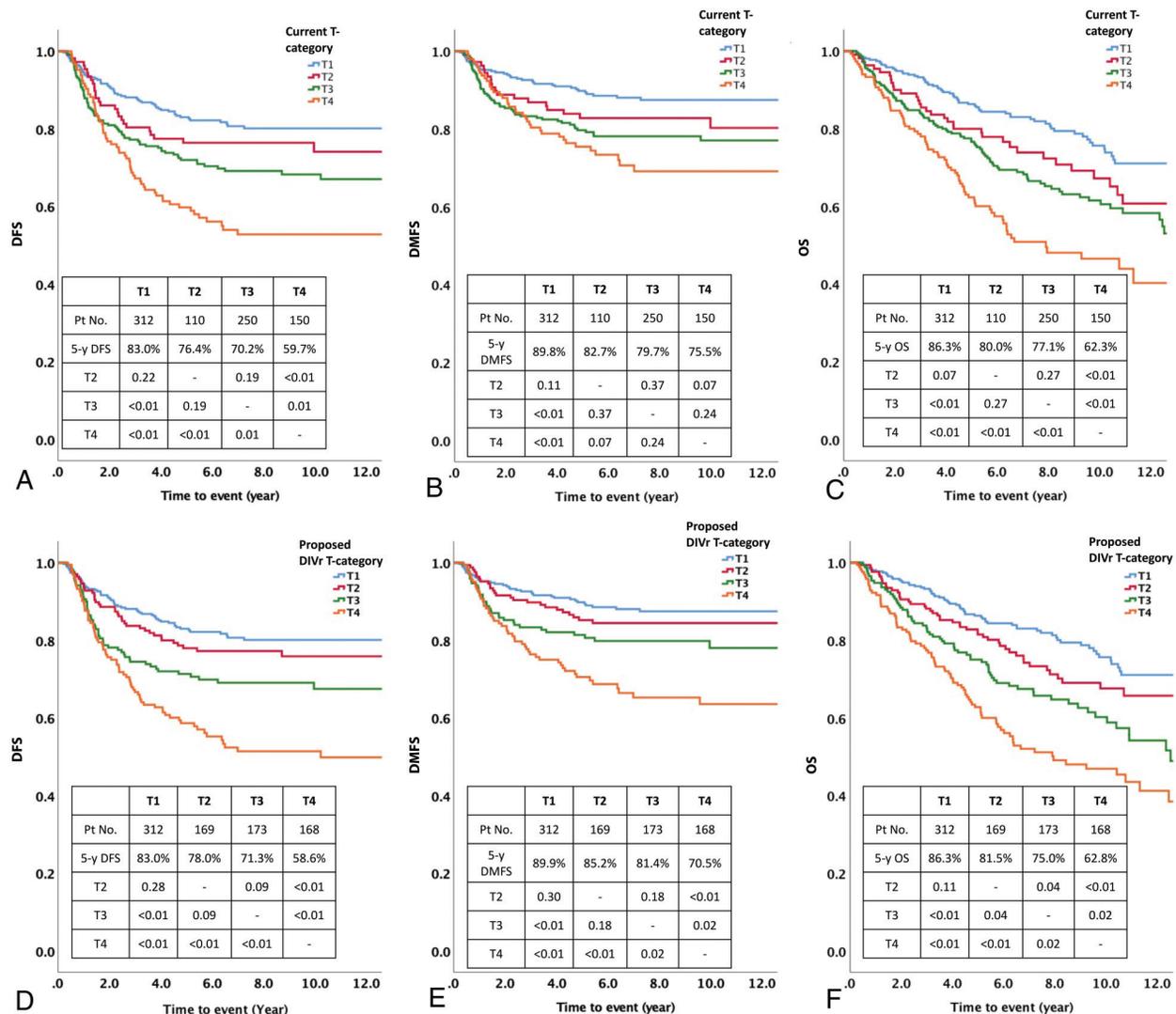
Bolded *P* values indicate statistical significance.

Note:—HR indicates hazard ratio.

**Table 3: C-index of the current AJCC T-category and DIVr for DFS, DMFS, and OS in the duration from baseline to 3-year, 5-year, and overall follow-up**

	DFS		DMFS		OS	
	C-Index (95% CI)	P Value	C-Index (95% CI)	P Value	C-Index (95% CI)	P Value
Three-year						
T-category	0.608 (0.606–0.609)	Ref.	0.602 (0.600–0.604)	Ref.	0.624 (0.622–0.626)	Ref.
DIVr	0.625 (0.623–0.627)	.07	0.634 (0.632–0.636)	<b>&lt;.01</b>	0.643 (0.641–0.645)	.12
Five-year						
T-category	0.599 (0.597–0.600)	Ref.	0.596 (0.595–0.597)	Ref.	0.617 (0.616–0.618)	Ref.
DIVr	0.615 (0.614–0.616)	.05	0.624 (0.622–0.626)	<b>&lt;.01</b>	0.629 (0.627–0.631)	.19
Overall						
T-category	0.602 (0.601–0.603)	Ref.	0.597 (0.595–0.599)	Ref.	0.598 (0.597–0.599)	Ref.
DIVr	0.620 (0.619–0.621)	<b>.03</b>	0.626 (0.625–0.627)	<b>&lt;.01</b>	0.608 (0.607–0.609)	.13

Bolded P values indicate statistical significance.



**FIG 2.** The Kaplan-Meier curves of the current (eighth edition) AJCC T-category (A–C) and the proposed DIVr-based T-category (D–F) for DFS (A and D), DMFS (B and E), and OS (C and F) in patients with NPC. Compared with the AJCC T-category, the proposed DIVr-based T-category only resulted in slight changes in the survival rates and numbers of patients in each of the T-categories.

of DIVr-T4 from DIVr-T3,  $P = .02$ , DIVr-T4 from DIVr-T2,  $P < .01$ ), which is not observed for AJCC-T4 category (5-year DMFS for AJCC-T4 = 75.5%, logrank test of AJCC-T4 from AJCC-T3,  $P = .24$ , AJCC-T4 from AJCC-T2,  $P = .07$ ).

**PTV, DIV, and DIVr Association with Nodal Metastases**

Univariable logistic regression showed that large DIVr (OR = 1.938, 95% CI = 1.269–2.960,  $P < .01$ ), DIV (OR = 1.012, 95% CI = 1.004–1.021,  $P < .01$ ), and PTV (OR = 1.014, 95% CI =

1.006–1.022,  $P < .01$ ) correlated with the presence of nodal metastases ( $n+$ ). The AUCs of DIV<sub>r</sub>, DIV, and PTV were 0.565, 0.574, and 0.604, respectively, for the prediction of  $n+$ , and the AUCs among these measurements had no statistical differences ( $P > .05$ ).

## DISCUSSION

This NPC study was performed in a large group of patients with long-term follow-up and is one of the first to investigate the association between the depth of primary tumor invasion and outcome after treatment with curative intent in this cancer. The results showed that high measurements for deep invasion of the primary NPC, expressed both as a volume (DIV) and as a ratio of the total tumor volume (DIV<sub>r</sub>), together with high total PTV, were predictors of poor outcome by using all 3 end points of DFS, DMFS, and OS. Of these 3 measurements the DIV<sub>r</sub> showed the strongest correlation with outcome. This is probably because compared with the DIV and PTV, the DIV<sub>r</sub> has a greater ability to capture the aggressive invasive pattern of tumors that preferentially invade deeply rather than superficially. In essence, tumors with poorer outcome were those where the deep component was relatively large compared with the superficial component. Many of these tumors had only a small component within the nasopharynx while the bulk of the tumor was invading deeply beyond the boundary of the nasopharynx into the adjacent tissues, including the parapharynx and retropharynx, skull base, and intracranial structures. The invasive edge of the tumor is an important predictor in patient outcome<sup>20,21</sup> and at the cellular level, signaling pathways such as ERK and PI3K signal transmission<sup>22,23</sup> mediate the invasion beyond the basement membrane, eventually leading to relatively uninhibited deep spread of tumor including into those surrounding structures that are associated with adverse outcome for NPCs, such as the prevertebral space,<sup>24,25</sup> skull base, and intracranium.<sup>26,27</sup> The DIV<sub>r</sub> was also associated with the presence of nodal metastases, ie, tumors with a higher propensity for deep invasion were more likely to have nodal metastases, although in this aspect, DIV<sub>r</sub> did not outperform either DIV or PTV. However, this association is interesting because it links features of the primary tumor to the development of nodal metastases and in turn nodal metastases have strong links to the development of distant metastases, especially nodal volume.<sup>28–30</sup>

We also evaluated by using DIV<sub>r</sub> thresholds for clinical use. A threshold of 0 represented superficial tumors without deep invasion, which is equivalent to the eighth edition AJCC-T1. The DIV<sub>r</sub> thresholds for deep invasion were divided into 3 ascending groups ( $>0\%$ ,  $\leq 60.0\%$ ;  $>60.0\%$ ,  $\leq 80.0\%$ ; and  $>80.0\%$ ) which reflected a similar structure to the current AJCC T-categories (T2, T3, T4). The thresholds for DIV<sub>r</sub> showed a slight improvement in the performance compared with the current AJCC T-categories while controlling for nodal status, use of chemotherapy, age, and sex. Using DIV<sub>r</sub> thresholds instead of the current AJCC T-categories showed a modest improvement in predicting DMFS and DFS but not OS. This was most significant for DMFS, suggesting that DIV<sub>r</sub> may be useful in identifying those patients at higher risk of developing distant metastases so that they may undergo more intensive posttreatment surveillance. Currently the

process to obtain these measurements is time-consuming but this could be facilitated in the future by artificial intelligence–based NPC tumor segmentation.<sup>8,9,12</sup>

This was the first study to evaluate DIV in NPC, but our results were in accordance with other carcinomas, such as oral cavity and rectal carcinoma where the extent of deep invasion reliably predicts treatment outcome.<sup>1,31</sup> For oral cavity cancers, the pathologic DOI is recognized as a major contributor to outcome and is included in the T-category for staging while unidimensional imaging measurements are used preoperatively for treatment planning. Similar to our study, the thresholds in oral cancer are divided into 3 groups to indicate nodal metastases and predict outcome.<sup>1,32</sup>

The NPC primary tumor volume has been shown to correlate with relapse and outcome, however, reported PTV thresholds show a wide range (13–50 mL)<sup>33–35</sup> and PTV is a weaker predictor than nodal volume. Our results for PTV showed modest correlation with outcome that is similar to those reported by Qin et al.<sup>36</sup>

We believe the study has several strengths. The long follow-up period was sufficient to capture most patients with NPC recurrence that typically happens within 5 years after treatment.<sup>37</sup> Our cohort with both early- and late-staged disease were representative of patients with nonmetastatic NPC at initial diagnosis and treatment of patients with early and locally advanced disease, and there was an excellent interobserver agreement on tumor segmentation. However, this study has some limitations. The recruitment was from a single center with patients of mostly Chinese population, whereas further validation of these results may be required in alternative ethnicities. MRI staging was performed after the NPC had been histologically confirmed by biopsy, which is our standard practice. We believe that the reduction in the volume of the superficial component of the tumor would have a negligible impact on the results, as the size of the biopsy specimens is usually small (2–3 mm in diameter). The volume measurements including DIV<sub>r</sub> were not correlated with levels of the plasma Epstein-Barr virus DNA because this test was not performed routinely in all our patients with NPC; furthermore although this test is extremely promising, it is not used in international staging classifications.<sup>38</sup>

## CONCLUSIONS

The DIV<sub>r</sub> reflects the propensity of NPC for deep invasion and was associated with nodal metastases on initial presentation and predicted survival especially in relationship to DMFS. This measurement performed better than PTV and improved slightly the performance of the AJCC T-category for DFS and DMFS but not OS. The additional value of using DIV<sub>r</sub> was not strong enough to justify incorporation into T-categorization for staging NPC but our results suggest it could have a role in guiding posttreatment imaging strategies for early detection of distant metastases.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Ebrahimi A, Gil Z, Amit M, et al; International Consortium for Outcome Research (ICOR) in Head and Neck Cancer. **Primary tumor staging for oral cancer and a proposed modification incorporating depth of invasion: an international multicenter retrospective study.** *JAMA Otolaryngol Head Neck Surg* 2014;140:1138–48 CrossRef Medline
2. Kuan EC, Clair JMS, Badran KW, et al. **How does depth of invasion influence the decision to do a neck dissection in clinically N0 oral cavity cancer?** *Laryngoscope* 2016;126:547–48 CrossRef Medline
3. Aaboubout Y, van der Toom QM, de Ridder MAJ, et al. **Is the depth of invasion a marker for elective neck dissection in early oral squamous cell carcinoma?** *Front Oncol* 2021;11:628320 CrossRef Medline
4. Amin MB, Edge SB, Greene FL, et al, eds. *AJCC Cancer Staging Manual*. 8th ed. Springer-Verlag; 2017.
5. Lydiatt WM, Patel SG, O'Sullivan B, et al. **Head and neck cancers—major changes in the American Joint Committee on cancer eighth edition cancer staging manual.** *CA Cancer J Clin* 2017;67:122–37 CrossRef Medline
6. Lee MK, Choi Y. **Correlation between radiologic depth of invasion and pathologic depth of invasion in oral cavity squamous cell carcinoma: a systematic review and meta-analysis.** *Oral Oncol* 2023;136:106249 CrossRef Medline
7. King AD, Zee B, Yuen EHY, et al. **Nasopharyngeal cancers: which method should be used to measure these irregularly shaped tumors on cross-sectional imaging?** *Int J Radiat Oncol Biol Phys* 2007;69:148–54 CrossRef Medline
8. Wong LM, Ai QYH, Poon DMC, et al. **A convolutional neural network combined with positional and textural attention for the fully automatic delineation of primary nasopharyngeal carcinoma on non-contrast-enhanced MRI.** *Quant Imaging Med Surg* 2021;11:3932–44 CrossRef Medline
9. Wong LM, Ai QYH, Mo FKF, et al. **Convolutional neural network in nasopharyngeal carcinoma: how good is automatic delineation for primary tumor on a non-contrast-enhanced fat-suppressed T2-weighted MRI?** *Jpn J Radiology* 2021;39:571–79 CrossRef Medline
10. Ke L, Deng Y, Xia W, et al. **Development of a self-constrained 3D DenseNet model in automatic detection and segmentation of nasopharyngeal carcinoma using magnetic resonance images.** *Oral Oncol* 2020;110:104862 CrossRef Medline
11. Wang Y, Chen H, Lin J, et al. **Automatic detection and recognition of nasopharynx gross tumour volume (GTVnx) by deep learning for nasopharyngeal cancer radiotherapy through magnetic resonance imaging.** *Radiat Oncol* 2023;18:76 CrossRef Medline
12. Ng WT, But B, Choi HCW, et al. **Application of artificial intelligence for nasopharyngeal carcinoma management – a systematic review.** *Cancer Manag Res* 2022;14:339–66 CrossRef Medline
13. King AD, Ahuja AT, Leung SF, et al. **Neck node metastases from nasopharyngeal carcinoma: MR imaging of patterns of disease.** *Head Neck* 2000;22:275–81 CrossRef Medline
14. van den Brekel M, Stel H, Castelijns J, et al. **Cervical lymph node metastasis: assessment of radiologic criteria.** *Radiology* 1990;177:379–84 CrossRef Medline
15. Brierley J, Gospodarowicz M, Wittekind C, eds. *UICC TNM Classification of Malignant Tumours*. 8th ed. Wiley-Blackwell Publishing Ltd; 2017:15
16. Harrell FE, Lee KL, Mark DB. **Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors.** *Statist Med* 1996;15:361–87 CrossRef
17. Uno H, Cai T, Pencina MJ, et al. **On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data.** *Stat Med* 2011;30:1105–17 CrossRef Medline
18. Correa AF, Jegede O, Haas NB, et al. **Predicting renal cancer recurrence: defining limitations of existing prognostic models with prospective trial-based validation.** *J Clin Oncol* 2019;37:2062–71 CrossRef Medline
19. Moons KGM, Altman DG, Reitsma JB, et al. **Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration.** *Ann Intern Med* 2015;162:W1–73 CrossRef Medline
20. Sandu K, Nisa L, Monnier P, et al. **Clinicobiological progression and prognosis of oral squamous cell carcinoma in relation to the tumor invasive front: impact on prognosis.** *Acta Otolaryngol* 2014;134:416–24 CrossRef Medline
21. Jimenez L, Jayakar SK, Ow TJ, et al. **Mechanisms of invasion in head and neck cancer.** *Arch Pathol Lab Med* 2015;139:1334–48 CrossRef Medline
22. O-Charoenrat P, Wongkajornsilp A, Rhys-Evans PH, et al. **Signaling pathways required for matrix metalloproteinase-9 induction by betacellulin in head-and-neck squamous carcinoma cells.** *Int J Cancer* 2004;111:174–83 CrossRef Medline
23. Curry JM, Sprandio J, Cognetti D, et al. **Tumor microenvironment in head and neck squamous cell carcinoma.** *Semin Oncol* 2014;41:217–34 CrossRef Medline
24. Feng AC, Wu MC, Tsai SYC, et al. **Prevertebral muscle involvement in nasopharyngeal carcinoma.** *Int J Radiat Oncol Biol Phys* 2006;65:1026–35 CrossRef Medline
25. Ai QY, Hu CW, Bhatia KS, et al. **Nasopharyngeal carcinoma: relationship between invasion of the prevertebral space and distant metastases.** *Eur Arch Otorhinolaryngol* 2018;275:497–505 CrossRef Medline
26. Cheng Y-K, Liu L-Z, Jiang N, et al. **MRI-detected skull-base invasion: prognostic value and therapeutic implication in intensity-modulated radiotherapy treatment for nasopharyngeal carcinoma.** *Strahlenther Onkol* 2014;190:905–11 CrossRef Medline
27. Chen L, Liu L, Mao Y, et al. **Grading of MRI-detected skull-base invasion in nasopharyngeal carcinoma and its prognostic value.** *Head Neck* 2011;33:1309–14 CrossRef Medline
28. Chen F-P, Zhou G-Q, Qi Z-Y, et al. **Prognostic value of cervical nodal tumor volume in nasopharyngeal carcinoma: analysis of 1230 patients with positive cervical nodal metastasis.** *PLoS One* 2017;12:e0176995 CrossRef Medline
29. Yuan H, Ai QY, Kwong DLW, et al. **Cervical nodal volume for prognostication and risk stratification of patients with nasopharyngeal carcinoma, and implications on the TNM-staging system.** *Sci Rep* 2017;7:10387 CrossRef Medline
30. Ai Q-Y, King AD, Mo FKF, et al. **Prediction of distant metastases from nasopharyngeal carcinoma: improved diagnostic performance of MRI using nodal volume in N1 and N2 stage disease.** *Oral Oncol* 2017;69:74–79 CrossRef Medline
31. Miyoshi M, Ueno H, Hashiguchi Y, et al. **Extent of mesorectal tumor invasion as a prognostic factor after curative surgery for T3 rectal cancer patients.** *Ann Surg* 2006;243:492–98 CrossRef Medline
32. Mukoyama N, Suzuki H, Hanai N, et al. **Pathological tumor volume predicts survival outcomes in oral squamous cell carcinoma.** *Oncol Lett* 2018;16:2471–77 CrossRef Medline
33. Chen C, Fei Z, Pan J, et al. **Significance of primary tumor volume and T-stage on prognosis in nasopharyngeal carcinoma treated with intensity-modulated radiation therapy.** *Jpn J Clin Oncol* 2011;41:537–42 CrossRef Medline
34. Lee C, Huang T, Lee M, et al. **Clinical application of tumor volume in advanced nasopharyngeal carcinoma to predict outcome.** *Radiat Oncol* 2010;5:2–7 CrossRef
35. Tian Y-M, Xiao W-W, Bai L, et al. **Impact of primary tumor volume and location on the prognosis of patients with locally recurrent nasopharyngeal carcinoma.** *Chin J Cancer* 2015;34:247–53 CrossRef Medline
36. Qin L, Wu F, Lu H, et al. **Tumor volume predicts survival rate of advanced nasopharyngeal carcinoma treated with concurrent chemoradiotherapy.** *Otolaryngol Head Neck Surg* 2016;155:598–605 CrossRef Medline
37. Lee AWM, Foo W, Law SCK, et al. **Recurrent nasopharyngeal carcinoma: the puzzles of long latency.** *Int J Radiat Oncol Biol Phys* 1999;44:149–56 CrossRef Medline
38. Wong ECY, Hung JLC, Ng WT. **Potential pitfalls in incorporating plasma Epstein-Barr virus DNA in the management of nasopharyngeal carcinoma.** *Head Neck* 2020;42:446–55 CrossRef Medline

# Peritumoral Hyperintense Signal on Postcontrast FLAIR Images Surrounding Vestibular Schwannomas Following Stereotactic Radiosurgery

 Sandy T. Nguyen,  John C. Benson,  Girish Bathla,  Paul J. Farnsworth,  Matthew L. Carlson,  Michael J. Link, and  John I. Lane

## ABSTRACT

**BACKGROUND AND PURPOSE:** Prior investigations have noted the presence of peritumoral hyperintense signal (a “halo”) around vestibular schwannomas on postcontrast 3D T2 FLAIR images. This study evaluated this phenomenon in a cohort of patients undergoing stereotactic radiosurgery.

**MATERIALS AND METHODS:** A retrospective review was completed of consecutive patients with presumed vestibular schwannomas undergoing stereotactic radiosurgery. Tumor size, location, presence or absence of a peritumoral halo, and halo thickness were recorded. Images were reviewed for presence and size of peritumoral hyperintense signal on postcontrast 3D T2 FLAIR images before and after treatment.

**RESULTS:** Twenty-six patients were included in this study, 14 of whom were women (54.0%). Average age was  $62 \pm 12$  years. Before treatment, a postcontrast 3D T2 FLAIR hyperintense peritumoral halo was seen in 85% of patients, averaging  $0.8 \pm 0.4$  mm in thickness. There was a higher incidence of peritumoral halo in posttreatment patients (96%) than pretreatment patients (85%) ( $P = .017$ ) with a mean follow-up period of 1.2 years (SD 0.35) from November 12, 2019, to September 5, 2023. The average halo thickness was also larger in posttreatment patients (average  $=1.4 \pm 0.4$  mm) compared with pretreatment patients ( $0.8 \pm 0.4$  mm) ( $P < .001$ ). Average tumoral size did not significantly change following treatment ( $P = .10$ ).

**CONCLUSIONS:** Vestibular schwannomas treated with stereotactic radiosurgery are more likely to have a peritumoral halo on postcontrast 3D T2 FLAIR images, with larger halo size as compared with pretreatment studies. Further study with a larger tumor cohort and longer follow-up will be necessary to determine if these findings are predictive of subsequent tumor shrinkage.

**ABBREVIATIONS:** CPA = cerebellopontine angle; IAC = internal auditory canal; SRS = stereotactic radiosurgery; VS = vestibular schwannoma

Vestibular schwannomas (VSs) are benign nerve sheath tumors that account for 6%–8% of all intracranial tumors and represent the most common tumor of the cerebellopontine angle (CPA). VSs classically arise within the internal auditory canal (IAC) and extend into the CPA with tumor growth.<sup>1,2</sup> On imaging, treatment naïve VSs can either demonstrate homogeneous or heterogeneous intratumoral enhancement and may contain hemorrhage or cystic changes. Edema of the adjacent parenchyma is seen in up to 40% of cases, particularly with large tumor size or rapid growth.<sup>3,4</sup> Symptoms of VSs most commonly encompass ipsilateral hearing loss and tinnitus. With increasing size, tumors may exert mass effect on adjacent

structures, resulting in trigeminal symptoms, hydrocephalus, and ataxia.<sup>5</sup>

A recent study described the presence of peritumoral hyperintense signal or “halo” around VSs on postcontrast 3D T2 FLAIR images. The authors of that study opined that this halo represented local leakage of gadolinium into the peritumoral space, though the mechanism remains unknown.<sup>9</sup> To date, however, investigation of this phenomenon has been limited and restricted to treatment-naïve patients. As such, the current study sought to further our knowledge of this subject by comparing the incidence and thickness of peritumoral halos in patients undergoing stereotactic radiosurgery (SRS).

Received May 15, 2024; accepted after revision January 9, 2025.

From the Departments of Radiology (S.T.N., J.C.B., G.B., P.J.F., J.I.L.), Otolaryngology-Head and Neck Surgery (M.L.C., M.J.L.), and Neurologic Surgery (M.L.C., M.J.L.), Mayo Clinic, Rochester, Minnesota.

Please address correspondence to John C. Benson, Department of Radiology, Mayo Clinic, 200 1st St SW, Rochester, MN 55905; e-mail: Benson.john3@mayo.edu  
<http://dx.doi.org/10.3174/ajnr.A8657>

## MATERIALS AND METHODS

### Eligibility Criteria

Local institutional review board approval was received before the study commencement. A retrospective review of patients with clinically presumed VSs was conducted based on imaging

## SUMMARY

**PREVIOUS LITERATURE:** This is a novel investigation of peritumoral hyperintense signal (a “halo”) around vestibular schwannomas in a patient cohort undergoing stereotactic radiosurgery.

**KEY FINDINGS:** Treated vestibular schwannomas tend to have great incidence of peritumoral “halo” with increased thickness compared with pretreatment lesions.

**KNOWLEDGE ADVANCEMENT:** Increased understanding of signal changes of vestibular schwannomas following SRS could relate to increased tumor permeability and may be a predictor of subsequent tumor shrinkage.

findings over 4 years (11/12/2019–8/23/2023). These tumors were treatment-naïve and subsequently underwent SRS. Treatment with SRS was planned with the 3D T1 postcontrast sequence coregistered to the contour of the 50% isodose line. Patients with dedicated IAC imaging before and following therapy were reviewed for appropriate pre- and postcontrast images for comparison. Patients were excluded for absence of necessary MR imaging sequences. Demographic information was acquired through the electronic medical record.

### MR Imaging Protocol

Patients were scanned with either a 3T Siemens or GE Healthcare scanner with multichannel phased array coils (and a 32- or 64-channel head coil). Twenty-three of the 26 patients underwent evaluation with a Siemens scanner both before and following SRS. One patient before SRS and 2 patients following SRS were imaged via a GE Healthcare scanner. Dedicated IAC imaging was performed utilizing axial 3D T1 sampling perfection with application-optimized contrasts by using different flip angle evolution (SPACE sequence; Siemens) (TR = 600 ms, TE = 32 ms, data matrix = 192 × 192, acquisition time = 4 minutes), axial 3D T2 SPACE (TR = 1300 ms, TE = 184 ms, data matrix = 320 × 320, acquisition time = 3 minutes 55 seconds), axial 3D fat-saturated postcontrast T1 SPACE (TR = 600 ms, TE = 32 ms, data matrix = 192 × 192, acquisition time = 4 minutes), and axial postcontrast 3D T2 FLAIR (TR = 5000 ms, TE = 379 ms, data matrix = 192 × 192, inversion time = 1700 ms, acquisition time = 4 minutes 29 seconds). The FOV was 150 for these sequences.<sup>10</sup> Images were acquired in the sequence noted above per institutional protocol and also included precontrast standard full brain sequences (ie, axial fat-saturated precontrast T2 FLAIR and sagittal precontrast T1 FLAIR). Gadobutrol (Gadavist) per weight-based table (0.1 mmol/kg) was administered via IV push for all the examinations. Axial 3D fat-saturated postcontrast T1 SPACE and, subsequently, axial postcontrast 3D T2 FLAIR were acquired following contrast injection. The order of the sequences was kept the same for all performed studies.

### Imaging Evaluation

Two experienced neuroradiologists (J.C.B., J.I.L.) performed the retrospective imaging review. The radiologists were blinded to the presence of recent treatment. VSs were evaluated for laterality—right or left—and location, described as in the IAC, CPA, or both. Before and following treatment, all VSs were measured in a single maximal axial dimension for size. If the VS involved

both the IAC and CPA, the largest single axial dimension of the CPA component was measured for standardization. Presence or absence of a fundal cleft was also reported, as well as a fundal cleft size between the lateral aspect of the VS and the IAC fundus.

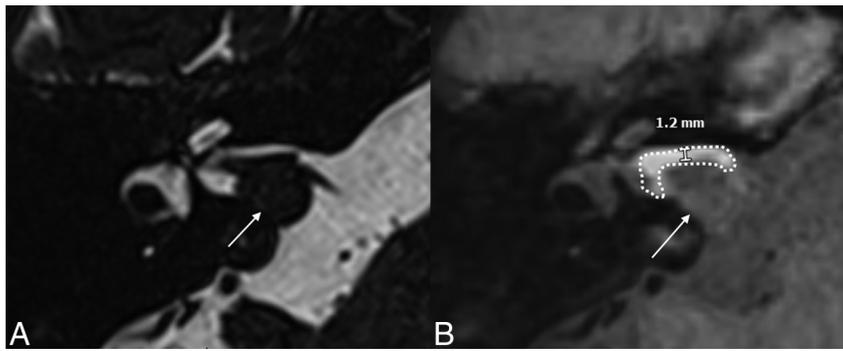
VSs were evaluated for presence or absence of a peritumoral hyperintense signal on postcontrast 3D T2 FLAIR images (“halo”), before and after SRS, through the overlaying and fusion features of our PACS system (Visage, Version 7.1.18, Visage Imaging). Axial postcontrast 3D T2 FLAIR images were fused and overlaid with both axial fat-saturated postcontrast T1W1 and axial T2 SPACE images (Fig 1). Precontrast T2 FLAIR and postcontrast 3D T2 FLAIR were not directly comparable, and therefore, the halos could not be measured on precontrast images. Size or thickness of halo, if present, was measured through the differences in tumor size between the different sequences. The halos were measured perpendicular to the tumors. Though the halos were irregular in areas, the largest halo measurement for each tumor was recorded. Image subtraction was also utilized to better detect differences between the tumor margins and the peritumoral halo. Whether the peritumoral hyperintense signal was confined to the fundus was recorded. Halo thickness was averaged between the 2 observer measurements. Any disagreements between the observers for categorical variables were resolved by consensus.

### Statistical Analysis

For all continuous variables, means and SDs were calculated. Pearson  $\chi^2$  test was used to calculate statistically significant associations between the continuous variables. The Student *t* test was used to calculate statistical differences of the continuous variables. Cohen  $\kappa$  was used to calculate interrater reliability with categorical variables. *P* values <.05 were statistically significant.

## RESULTS

Twenty-six patients were included in the study, of which 14 were women (54.0%) and 12 were men (46.0%). Average age was 62 years (SD 12). Fourteen tumors were right-sided (56.0%), and 12 were on the left (46%). Of these, 8 (31%) tumors were restricted to the IAC, 1 (4%) was entirely situated in the cerebello-pontine angle, and 17 (65%) involved both regions. The tumors were Koos 1–4: 7 Koos 1 (27%), 11 Koos 2 (42%), 5 Koos 3 (19%), and 3 Koos 4 (12%). Average follow up period was 1.2 (SD 0.35) years or, on average, 14 months for each patient following initial baseline imaging. Pretreatment MRI studies were performed



**FIG 1.** Evaluation and measurement of peritumoral halo. Axial T2 SPACE (A) and coregistered T2 SPACE and postcontrast 3D T2 FLAIR (B) images demonstrate a right-sided vestibular schwannoma (arrows) with peritumoral enhancement that is predominantly along the anterior margin of the tumor (dashed line). Measurement of the halo was performed perpendicular to the tumoral margin.

immediately before treatment. Postprocedural imaging was acquired approximately 10 (SD 2.4) months following end of treatment. Pretreatment VSs measured  $1.1 \pm 0.4$  cm, and post-treatment VSs measured  $1.3 \pm 3.3$  cm. Average tumoral size did not significantly change following treatment ( $P = .10$ ). Interobserver agreement on the presence of a halo before treatment was substantial ( $\kappa = 0.63$ ), whereas following treatment, it was moderate ( $\kappa = 0.56$ ).

Peritumoral halo on postcontrast 3D T2 FLAIR images (“halo”) was seen in 85% of patients before treatment and 96% following SRS, with significant difference noted between pre- and posttreatment groups ( $P = .017$ ). Before treatment, peritumoral halos measured  $0.8 \pm 0.4$  mm in maximum thickness on average, which significantly increased in size following treatment ( $1.4 \pm 0.4$  mm) ( $P < .001$ ). Four of the VSs did not demonstrate a peritumoral halo before SRS (15.0%). Of these 4, 3 developed halos following treatment.

A fundal cleft was observed in 21 (81%) patients, with an average cleft size of  $1.7 \pm 1.1$  mm. The cleft size did not significantly change following treatment ( $P = .14$ ). Peritumoral halo was restricted to the fundus in 18 (69.0%) patients before SRS. None of the posttreatment patients had a halo restricted to the fundus. There was no correlation between the change in halo thickness and Koos classification.

## DISCUSSION

This study evaluated the changes in the incidence and thickness of peritumoral halos around VSs following SRS in a treatment-naïve cohort. The results indicate that peritumoral halos are more commonly seen around posttreatment tumors and are increased in thickness. The halos, when present, are also less likely to be restricted to the IAC fundus. Together, these findings indicate that the incidence, thickness, and locality of peritumoral halos around VSs change after SRS.

Stereotactic radiosurgery treatment combines delayed vascular and cytotoxic effects as described by Yang et al.<sup>11</sup> These delayed vascular effects pertain to radiation-induced damage to tumor nutrient vessels, critical for tumoral necrosis. The underlying abnormal vasculature of VSs is more susceptible to radiation-

induced sequelae than normal vessels. Cytotoxic effects relate to DNA damage from  $\gamma$  rays, generating oxygen free radicals and resultant strand breakage. The effects of SRS were studied on healthy rat brains, generating substantial vascular permeability about the area of treatment and the formation of new, leaky blood vessels seen on Gd-DTPA dynamic contrast-enhanced MRI and T2\*-weighted sequences.<sup>12</sup>

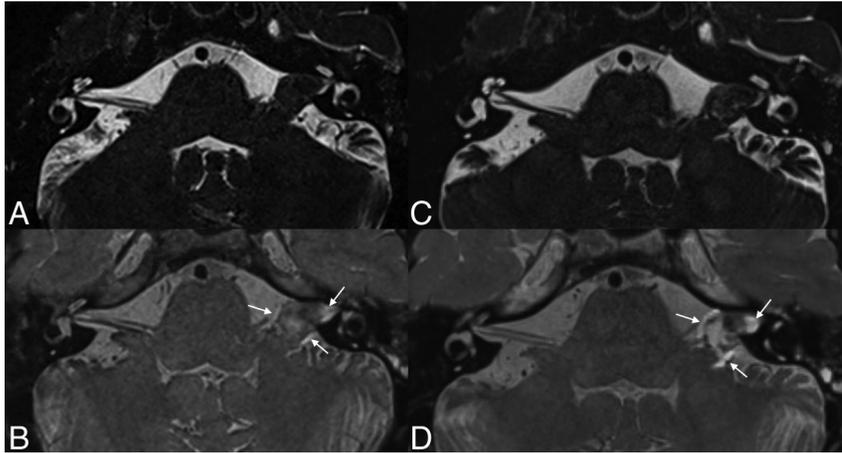
The most immediate implication of this study is that it offers further clues regarding the composition of peritumoral halos. Like the authors of the Benson et al<sup>9</sup> study, we hypothesize that the peritumoral halos could represent

local extra-tumoral leakage of gadolinium. The observed changes in these halos following SRS might represent a greater degree of gadolinium leakage after SRS. Though the findings might reflect increased tumoral permeability, it is important to mention that no additional references have theorized increased vascular permeability as the etiology of the halo sign, and the appearance could also partly be due to accompanying inflammation. If that is the case, steroid use after treatment might have affected our results.

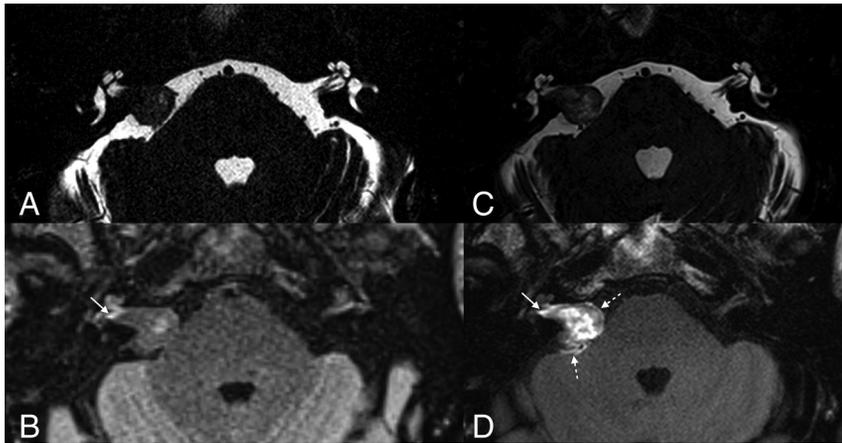
Steroids are not routinely prescribed in the early period after treatment. Though if a patient experiences a sudden sensorineural hearing loss in the post-SRS period (confirmed on audiogram), a course of steroids can be considered. In the uncommon case of a patient experiencing symptomatic brainstem edema after radiosurgery, steroids are used. In our cohort, no known steroid administration is documented to mitigate an inflammatory treatment response.

As first described, the peritumoral halos are often nonuniform in appearance and irregularly marginated. The authors of the original study on this concept hypothesized this irregularity could be explained by the trapping of gadolinium in the variably adhered and irregular arachnoid about the tumor.<sup>9</sup> It is possible that SRS leads to increased permeability and a larger concentration gradient of extravasated contrast, resulting in larger signal thickness. Extension of the peritumoral halo beyond the IAC might also be related to the internal necrosis of the tumor making it more “leaky.” Post-SRS tumor necrosis is an important correlate of treatment effect and might have a correlation with the degree of increased peritumoral halo. Evaluation of this relationship was not within the scope of this study but is worthy of future analysis.

Next, as stated in the results, peritumoral halos were often restricted to the IAC in pretreatment tumors but were observed circumferentially around the tumors in posttreatment patients. Most likely, this is because local gadolinium leakage is easier to observe in the IAC in the presence of a VS, where it is entrapped, or at least partially entrapped, by mass effect related to the VS. In posttreatment tumors, conversely, the increased permeability allows for observation of a halo even in regions that are not



**FIG 2.** Increased size of a peritumoral halo in a 68-year-old woman. Preradiosurgery axial T2 SPACE (A) and postcontrast 3D T2 FLAIR (B) images demonstrate a small halo located circumferentially around a left VS (arrows on B). Postradiosurgery axial T2 SPACE (C) and postcontrast 3D T2 FLAIR (arrows, D) images show that the halo has increased in size following treatment.



**FIG 3.** Increased halo size following SRS in a 66-year-old man. Pre-SRS axial T2 SPACE (A) and postcontrast 3D T2 FLAIR (B) images demonstrate minimal halo that is restricted to the IAC fundus (solid arrow, B). Following treatment, axial T2 SPACE (C) and postcontrast 3D T2 FLAIR (D) images show that the involvement of the fundus has increased (solid arrow, D). There is also a new circumferential halo elsewhere (dashed arrows, D).

entrapped, such as the cerebellopontine angle and cerebellum (Figs 2 and 3).

It is interesting to speculate that, if indeed, the halo corresponds to gadolinium leakage, its increase following treatment could reflect increased tumor permeability and may be a predictor of subsequent tumor shrinkage following SRS. Though we found no significant change in size of tumor after SRS in our small cohort, our follow-up period was less than 2 years. A longer follow-up period (potentially at least 3 years, as seen in a study performed by Lipski et al<sup>13</sup>) may be needed to see treatment response as it relates to tumor control and shrinkage, despite increased halo thickness and possible associated tumor permeability presenting within 2 years following treatment. Further study with a larger number of tumors over a longer imaging interval will be needed to determine if this might prove to be the case.

This study has several limitations given its retrospective nature. The results are based on a small cohort, and a larger sample size is needed to validate these findings and the overall clinical significance. Longer follow-up periods than what is seen in our cohort may also be helpful to evaluate for presence or absence of communicating hydrocephalus, change in tumor size, and temporal evolution of halo thickness and morphology, as well as clinical impact relating to sensorineural hearing loss, tinnitus, and vertigo. Unlike in our prior investigations, patients were selected for the availability of appropriate pre- and post-contrast imaging before and after treatment, which resulted in several patient exclusions for missing imaging. There are also inherent technical differences across multiple manufacturers, such as GE Healthcare and Siemens, which can be more challenging in direct comparison. Finally, the small average halo thickness of reported measurements—between 0.4 and 1.5 mm before treatment and 0.4 and 2.6 mm after SRS—may be challenging because of pixel size on standard PACS measuring tools. It is worth noting that the voxel dimension is 0.78 mm (192 × 192 acquisition matrix with FOV 150), and the average increase in halo thickness from pre- to posttreatment was 0.6 mm and, therefore, smaller than the acquired voxel.

## CONCLUSIONS

Vs treated with stereotactic radiosurgery are more likely to have a peritumoral halo on postcontrast 3D T2 FLAIR images, with larger halo sizes as

compared with pretreatment studies. Further study with a larger tumor cohort and longer follow-up will be necessary to determine if these findings are predictive of subsequent tumor shrinkage.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Lin EP, Crane BT. **The management and imaging of vestibular schwannomas.** *AJNR Am J Neuroradiol* 2017;38:2034–43 CrossRef Medline
2. Connor SE. **Imaging of the vestibular schwannoma: diagnosis, monitoring, and treatment planning.** *Neuroimaging Clin N Am* 2021; 31:451–71 CrossRef Medline
3. Bonneville F, Savatovsky J, Chiras J. **Imaging of cerebellopontine angle lesions: an update, Part 1: enhancing extra-axial lesions.** *Eur Radiology* 2007;17:2472–82 CrossRef Medline

4. Thamburaj K, Radhakrishnan VV, Thomas B, et al. **Intratumoral microhemorrhages on T2\*-weighted gradient-echo imaging helps differentiate vestibular schwannoma from meningioma.** *AJNR Am J Neuroradiol* 2008;29:552–57 CrossRef Medline
5. Gurewitz J, Schnurman Z, Nakamura A, et al. **Hearing loss and volumetric growth rate in untreated vestibular schwannoma.** *J Neurosurg* 2022;136:768–75 CrossRef Medline
6. Tanaka Y, Kobayashi S, Hongo K, et al. **Clinical and neuroimaging characteristics of hydrocephalus associated with vestibular schwannoma.** *J Neurosurg* 2003;98:1188–93 CrossRef Medline
7. Fukuda M, Oishi M, Kawaguchi T, et al. **Etiopathological factors related to hydrocephalus associated with vestibular schwannoma.** *Neurosurgery* 2007;61:1186–92 CrossRef Medline
8. Al Hinai Q, Zeitouni A, Sirhan D, et al. **Communicating hydrocephalus and vestibular schwannomas: etiology, treatment, and long-term follow-up.** *J Neurol Surg B Skull Base* 2013;74:68–74 CrossRef Medline
9. Benson JC, Carlson ML, Lane JI. **Peritumoral signal on postcontrast FLAIR images: description and proposed biomechanism in vestibular schwannomas.** *AJNR Am J Neuroradiol* 2023;44:1171–75 CrossRef Medline
10. Benson JC, Carlson ML, Lane JI. **MRI of the internal auditory canal, labyrinth, and middle ear: how we do it.** *Radiology* 2020;297:252–65 CrossRef Medline
11. Yang SY, Kim DG, Chung HT, et al. **Evaluation of tumour response after gamma knife radiosurgery for residual vestibular schwannomas based on MRI morphological features.** *J Neurol Neurosurg Psychiatry* 2008;79:431–36 CrossRef Medline
12. Constanzo J, Masson-Côté L, Tremblay L, et al. **Understanding the continuum of radionecrosis and vascular disorders in the brain following gamma knife irradiation: an MRI study.** *Magn Reson Med* 2017;78:1420–31 CrossRef Medline
13. Lipski SM, Hayashi M, Chernov M, et al. **Modern gamma knife radiosurgery of vestibular schwannomas: treatment concept, volumetric tumor response, and functional results.** *Neurosurg Rev* 2015;38:309–18; discussion 318 CrossRef Medline

# The Sensitivity of Arterial Spin-Labeling Imaging for Detection of Head and Neck Paragangliomas

Yannan Yu, Brian Q. Tsui, Christine M. Glastonbury, and Xin (Cynthia) Wu



## ABSTRACT

**BACKGROUND AND PURPOSE:** Head and neck paragangliomas (HNPGs) are rare neuroendocrine tumors whose hypervascular nature allows differentiation from many other head and neck neoplasms. We aimed to investigate the sensitivity of arterial spin-labeling (ASL) MR sequences for the detection of HNPGs.

**MATERIALS AND METHODS:** All head and neck MR examinations performed at a single tertiary institution between 2015 and 2023 were searched. Studies using ASL sequences that indicated either clinical suspicion for or ultimate imaging diagnosis of HNPG were identified. These studies were independently reviewed by 2 neuroradiologists blinded to the original radiology reports to determine, in a stepwise fashion, the following: 1) whether there was asymmetrically elevated blood flow on ASL imaging, 2) whether ASL findings correlated with lesions identifiable on conventional anatomic images, and 3) whether lesions likely reflected paragangliomas on the basis of correlations with clinical, laboratory, pathology, and other radiology data (Disagreement between raters was resolved by consensus.). The Cohen  $\kappa$  coefficient and the sensitivity of ASL in identifying HNPGs were calculated.

**RESULTS:** Eighty-four patients were included in the analysis (mean age, 54 [SD, 14] years and 47 women). Thirty patients had lesions confirmed or found likely to be HNPG, and 54 patients had lesions found unlikely to be HNPG or had no identifiable lesion. Among 46 of 84 patients with ASL blood flow asymmetry, 43 (93%) had lesions correlated with a lesion identifiable on anatomic imaging. Asymmetrically elevated ASL blood flow that correlated with a lesion demonstrated a sensitivity of 100% for reader A and 97% for reader B for identifying HNPG. The Cohen  $\kappa$  coefficient was 0.90 (SD, 0.11) between the 2 readers ( $P < .001$ ). Among 18 cases with pathology- or dotatate PET-proved HNPG, the sensitivity was 100% for reader A and 94% for reader B.

**CONCLUSIONS:** Asymmetrically elevated blood flow on ASL imaging demonstrates high sensitivity for the detection of HNPG, with almost perfect interrater agreement.

**ABBREVIATIONS:** ASL = arterial spin-labeling; HN = head and neck; HNPG = head and neck paraganglioma

Paragangliomas are rare hypervascular neuroendocrine tumors arising from sympathetic or parasympathetic paraganglia. Sympathetic paragangliomas typically secrete catecholamines and usually arise from the paravertebral ganglia of the chest, abdomen, and pelvis. Rarely, functional paragangliomas can occur in the head and neck (HN). In contrast, parasympathetic paragangliomas do not secrete catecholamines (ie, They are nonfunctional.). They are mostly located along the vagal and glossopharyngeal nerves in the HN, arising at the carotid body, nodose ganglion of the vagus nerve, jugular bulb, tympanic plexus, and larynx.<sup>1,2</sup>

Imaging is a crucial component of diagnostic assessment of paragangliomas. A <sup>68</sup>Ga/<sup>64</sup>Cu-dotatate scan is considered the most accurate imaging technique<sup>3,4</sup> in detecting paraganglioma, but the cost and availability of dotatate are variable.<sup>5</sup> Contrast-enhanced MRI and CT are more commonly used as the first-line imaging technique for suspected HN tumors, with MRI benefiting from higher soft-tissue contrast resolution in the neck.

Compared with most other HN masses, head and neck paraganglioma (HNPG) demonstrates hypervascularity and hyperperfusion. On CT, arterial or early venous phase hyperenhancement is a valuable distinguishing feature of HNPG. Similarly, studies have shown that MR TOF angiography and MR perfusion can differentiate paragangliomas and other HN masses such as meningiomas, schwannomas, and metastases.<sup>6-8</sup>

In contrast to conventional dynamic contrast-enhanced or dynamic-susceptibility contrast MR perfusion, both of which require IV gadolinium contrast, the arterial spin-labeling (ASL) technique is

Received October 29, 2024; accepted after revision January 11, 2025.

From the Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco California.

Please address correspondence to Xin Wu, MD, Department of Radiology & Biomedical Imaging, University of California San Francisco, 505 Parnassus Ave, San Francisco, CA 94143; e-mail: xin.wu@ucsf.edu; @CynXinWu



This article contains an [imaging protocol](#) available on our website.

<http://dx.doi.org/10.3174/ajnr.A8727>

## SUMMARY

**PREVIOUS LITERATURE:** HNPGs are rare neuroendocrine tumors that demonstrate hypervascularity and hyperperfusion. Previous studies have shown that MRA and dynamic contrast-enhanced MR perfusion can help differentiate between HNPGs and other HN masses. ASL, a perfusion method that is relatively resistant to susceptibility artifacts present in the HN region, may offer added diagnostic value in the evaluation of HNPGs without needing IV contrast, but its sensitivity for doing so remains underexplored.

**KEY FINDINGS:** Asymmetrically elevated ASL blood flow demonstrated 97%–100% sensitivity and 74% specificity for diagnosing HNPGs with a Cohen  $\kappa$  coefficient of 0.90, indicating almost perfect interrater reliability. ASL in addition to conventional MR sequences had a sensitivity of 100% and specificity of 93% in diagnosing HNPGs.

**KNOWLEDGE ADVANCEMENT:** This study provides preliminary data supporting the use of ASL in conventional head and neck MR protocols to better detect and increase the diagnostic confidence for HNPGs, given its high sensitivity and almost perfect interrater reliability for these lesions.

a noninvasive MR perfusion technique. ASL uses radiofrequency-labeled arterial protons in the entry arteries as contrast material and estimates the transit delay from the labeled vessel to the tissue of interest.<sup>9,10</sup> This transit delay provides a relatively accurate estimation of blood flow without the administration of contrast.<sup>11,12</sup> Due to its ease of implementation, ASL is used in many HN and brain protocols at our institution.

However, to our knowledge, the sensitivity of ASL to detect HNPGs has not been established. We aimed to investigate the interrater reliability, sensitivity, and specificity of ASL to detect HNPGs in patients with suspected and incidental paragangliomas.

## MATERIALS AND METHODS

A Health Insurance Portability and Accountability Act–compliant, institutional review board–approved search of all radiologic reports of MRIs of the HN (including the brain, skull base, internal auditory canals, face, and neck) dated between 2015 and 2023 from a single institution was conducted with the following search terms: “glomus” or “jugulare” or “jugulotympanicum” or “paraganglioma” or “carotid body.” These search terms were used to identify examinations that were ordered with the clinical indication for paraganglioma and those with an imaging diagnosis of paraganglioma, even if the diagnosis was not initially clinically suspected. MRI was obtained with 3T MR from multiple manufacturers and models, including Magnetom Vida (Siemens), Ingenia (Philips Healthcare), Signa Premier (GE Healthcare), Signa HDxt (GE Healthcare), and Discovery MR750 (GE Healthcare). Protocols included MR brain, MR internal auditory canal, MR face, and MR neck. ASL sequence parameters are as follows: 3D pseudocontinuous ASL with labeling time = 2000 ms, postlabeling delay = 2.0 seconds, and labeling plane offset = 2 cm; TE = 22 ms, TR = 4800 ms, inversion time = 3820 ms, number of averages = 1, flip angle = 120°, slice thickness = 4 mm, matrix size = 64 × 64, bandwidth = 3005 Hz/px.

From the initial search results, studies without ASL acquisitions were excluded. The remaining cases were independently reviewed by 2 neuroradiologists (reader A with 12 years of experience, reader B with 5 years of experience) blinded to the original radiology reports. The presence or absence of asymmetrically elevated ASL blood flow was recorded while the neuroradiologists

were blinded to other sequences and clinical information. “Asymmetric ASL blood flow” was defined as appreciable solid-appearing ASL blood flow asymmetry larger than the voxel size variation and higher than adjacent cerebral cortical blood flow. Conventional image sequences, including pre- and postcontrast T1-weighted imaging, T2 fat-suppressed imaging, or T2-weighted FLAIR and DWI, were then reviewed to determine whether ASL findings correlated with lesions identifiable on anatomic images. Finally, the readers made final diagnoses on whether the lesions likely reflected paragangliomas on the basis of all available clinical data, including the results of other imaging modalities such as dotatate PET scans and surgical pathology, when available. Disagreement between the 2 readers was resolved with a consensus review.

Relevant clinical information was obtained through chart review, including age, sex, MR study protocol, MR study indication, pathology and dotatate PET data, history of hereditary paraganglioma syndrome or high-risk mutation carrier status, and paraganglioma location and sidedness. Statistical analyses were conducted in STATA (Stata/MP 18; StataCorp). Continuous data were expressed as mean (SD), and categorical data were expressed as a counts with percentages. The Cohen  $\kappa$  coefficient was used to evaluate interrater reliability.<sup>13</sup> The  $\chi^2$  or Fisher exact test and independent-samples *t* test were used for comparisons of clinical data when appropriate. The sensitivity and specificity of identifying paraganglioma using asymmetrically elevated ASL blood flow were calculated in all patients and several subgroups, including patients with dotatate PET and/or pathology-confirmed paragangliomas, patients with pathology or other criterion standard tests proving nonparaganglioma, and studies with and without an indication of paraganglioma. We obtained the exact 95% CIs for sensitivity and specificity from the binomial distribution.

The methodology proposed by the Standards for Reporting of Diagnostic Accuracy Studies (STARD) checklist was followed.

## RESULTS

The initial search yielded studies from 193 patients, from which 109 (109/193, 56%) patients were excluded due to lack of the ASL acquisition. Eighty-four (84/193, 44%) patients were therefore included in the study, with a mean age of 54 (SD, 14) years, 47 (47/84, 56%) of whom were women.

**Table 1: Patient demographic and clinical data**

	Confirmed or Likely Paraganglioma, (n = 30)	Unlikely Paraganglioma or No Lesion, (n = 54)	P Value
Female, No. (%)	20 (67)	27 (50)	.14
Age (yr)	Mean, 52 (SD, 14)	Mean 55 (SD, 15)	.4
Study indication			
Suspect paraganglioma or surveillance of known paraganglioma, No. (%)	24 (80)	13 (24)	<.001
Pulsatile tinnitus, hearing loss, dizziness, or lightheadedness, No. (%)	3 (10)	22 (41)	
Evaluation of mass or other tumor, No. (%)	3 (10)	16 (30)	
Other, No. (%)	0 (0)	3 (5)	
Hereditary paraganglioma syndrome or high risk mutation, No. (%)	16 (53)	9 (17)	<.001
History of paraganglioma, No. (%)	25 (83)	6 (11)	<.001
Paraganglioma location <sup>a</sup>			
Jugular, No. (%)	14 (47)		
Tympanic, No. (%)	5 (17)		
Vagal, No. (%)	9 (30)		
Carotid body, No. (%)	12 (40)		
Other, No. (%)	2 (7)		

<sup>a</sup>Nine patients had paragangliomas in multiple locations.

**Table 2: Elevated ASL blood flow in detecting paraganglioma**

	Confirmed/Likely Paraganglioma, (n = 30)	Unlikely Paraganglioma or No Lesion, (n = 54)	P Value
Asymmetric ASL blood flow elevation correlated with a lesion			
Reader A, No. (%)	30 (100)	14 (26)	<.001
Reader B, No. (%)	29 (97)	14 (26)	<.001
Any asymmetric ASL blood flow elevation			
Reader A, No. (%)	30 (100)	16 (30)	<.001
Reader B, No. (%)	29 (97)	17 (31)	<.001
Likely paraganglioma on clinical radiology report, No. (%)	30 (100)	4 (7)	<.001

**Table 3: Sensitivity and specificity of ASL and clinical radiology reports**

	Sensitivity (%; 95% CI)	Specificity (%; 95% CI)
Asymmetric ASL blood flow elevation correlated with a lesion		
Reader A	100 (100–100)	74 (61–84)
Reader B	97 (79–100)	74 (61–84)
Any asymmetric ASL blood flow elevation		
Reader A, No. (%)	100 (100–100)	70 (57–81)
Reader B, No. (%)	97 (79–100)	69 (55–80)
Clinical radiology report at the time of the study	100 (100–100)	93 (89–100)

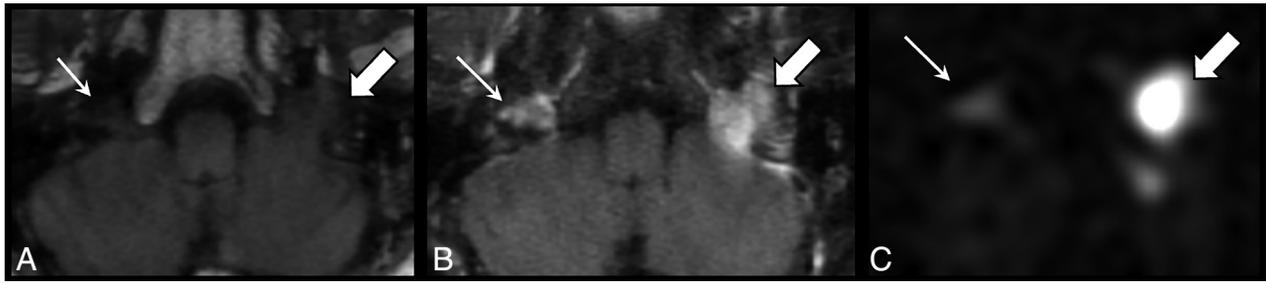
Thirty patients (30/84, 36%) had lesions confirmed or considered likely to be HNPG, of whom 15 (50%) had pathologic confirmation, and 3 (10%) had dotatate PET confirmation alongside an underlying high-risk genetic predisposition for paraganglioma formation. The remaining 12 (40%) patients in this cohort were considered likely to have HNPGs based on clinical and imaging information. Twenty-three patients (23/30, 77%) had unilateral HNPGs, and 7 (7/30, 23%) had bilateral lesions.

Twenty-five patients (25/84, 30%) had lesions considered unlikely to be paragangliomas, 12 of which (12/25, 48%) were proved nonparaganglioma by either pathology or other criterion standard examinations, including angiography-confirmed dural arteriovenous fistula ( $n=1$ ), pathology-proved hemangioma ( $n=1$ ), schwannoma ( $n=4$ ), meningioma ( $n=2$ ), epithelial neoplasm ( $n=2$ ), basal cell adenoma ( $n=1$ ), and osteomyelitis ( $n=1$ ). Twenty-nine patients (29/84, 35%) had no identifiable lesion. Demographic and clinical data are shown in Table 1.

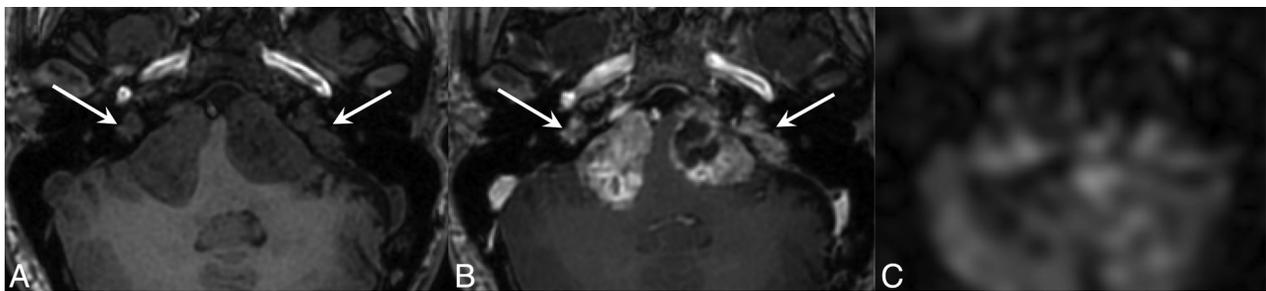
Both readers identified asymmetrically elevated ASL blood flow in 46 of the 84 cases (46/84, 55%). Of these 46 cases, reader A found that 44 (44/46, 96%) and reader B found that 43 cases

(43/46, 94%) had a correlation with lesions on anatomic imaging. The Cohen  $\kappa$  coefficient was 0.90 (SD, 0.11) ( $P < .001$ ) between the 2 readers for any asymmetrically elevated ASL blood flow and 0.88 (SD, 0.11) ( $P < .001$ ) for asymmetrically elevated ASL blood flow that correlated with a lesion, accepted to reflect almost perfect agreement.<sup>13</sup> There were 4 cases in which readers disagreed on ASL asymmetry, but they both agreed independently on whether the lesions were likely paragangliomas. The single case on which the 2 readers disagreed on the diagnosis was a subcentimeter right skull base lesion with associated ASL blood flow elevation and enhancement in a patient who had pathologically proved meningioma with hyperperfusion. The disagreement was resolved by consensus, and both readers agreed that this was not likely a paraganglioma. Unfortunately, there was no pathologic data to further characterize the lesion.

Asymmetrically elevated blood flow on ASL imaging that correlated with a lesion demonstrated a sensitivity of 100% (95% CI, 100%–100%) and specificity of 74% (95% CI, 61%–84%) for reader A and 97% (95% CI, 79%–100%) and 74% (95% CI, 61%–84%) for reader B for identifying paragangliomas (Tables 2 and 3).



**FIG 1.** Pathology-proved left jugular paraganglioma with ASL blood flow elevation. A female patient in her 50s initially presented with left-sided pulsatile tinnitus and was found to have an *SDHD* mutation. Surgical resection of a left middle ear mass revealed paraganglioma. Preoperative T1WI precontrast (A) and T1WI postcontrast (B) MR demonstrated an enhancing mass in the left jugular foramen extending into the left middle ear (*thick arrow*) and a smaller right jugular foramen mass (*thin arrow*). C, ASL blood flow map shows marked elevated blood flow associated with left jugular mass (*thick arrow*) and mildly elevated blood flow associated with right jugular mass (*thin arrow*). Surgical resection of the left middle ear portion of the mass revealed paraganglioma. The right-sided jugular foramen mass was not biopsied but was assumed to also represent paraganglioma, given the patient's underlying genetic mutation.



**FIG 2.** Pathology-proved nonparaganglioma (schwannoma) without ASL blood flow elevation. A female patient in her 30s with a history of neurofibromatosis type 2 with bilateral cerebellopontine angle masses. The patient underwent subtotal resection of a left-sided mass, which revealed a schwannoma, followed by radiation therapy. Preoperative axial T1WI precontrast (A) and T1WI postcontrast (B) MR shows bilateral cerebellopontine angle heterogeneously enhancing masses (*white arrows*). C, While mild signal heterogeneity was observed at the jugular foramen, the ASL blood flow map shows no significant solid blood flow elevation associated with the masses.

Among 18 cases with pathology- or dotatate-proved paraganglioma, the sensitivity was 100% (18/18; 95% CI, 100%–100%) for reader A and 94% (17/18; 95% CI, 65%–99%) for reader B. Among 15 cases with pathology-proved paraganglioma, the sensitivity was 100% (15/15; 95% CI, 100%–100%) for reader A and 93% (14/15; 95% CI, 58%–99%) for reader B. Among 12 cases with proved nonparaganglioma by pathology or other criterion standard examinations, reader A found 7 (58%) cases and reader B found 8 (67%) cases with asymmetric ASL blood flow. Cases with no lesions, only 1–2 cases (4%–7%, readers A and B, respectively) were rated as having asymmetrically-elevated ASL blood flow. Examples of true-positive, true-negative, and false-positive cases are shown in Figs 1–3.

For studies with suspected paraganglioma or surveillance of known paraganglioma history, reader A had a sensitivity of 100% (24/24; 95% CI, 100%–100%) and specificity of 92% (12/13; 95% CI, 59%–99%) and reader B had a sensitivity of 96% (23/24; 95% CI, 74%–99%) and specificity of 85% (11/13; 95% CI, 54%–96%) in this selected cohort. For studies with all other indications, both reader A and reader B detected asymmetrically elevated ASL blood flow in all incidental paragangliomas (6/6, 100%).

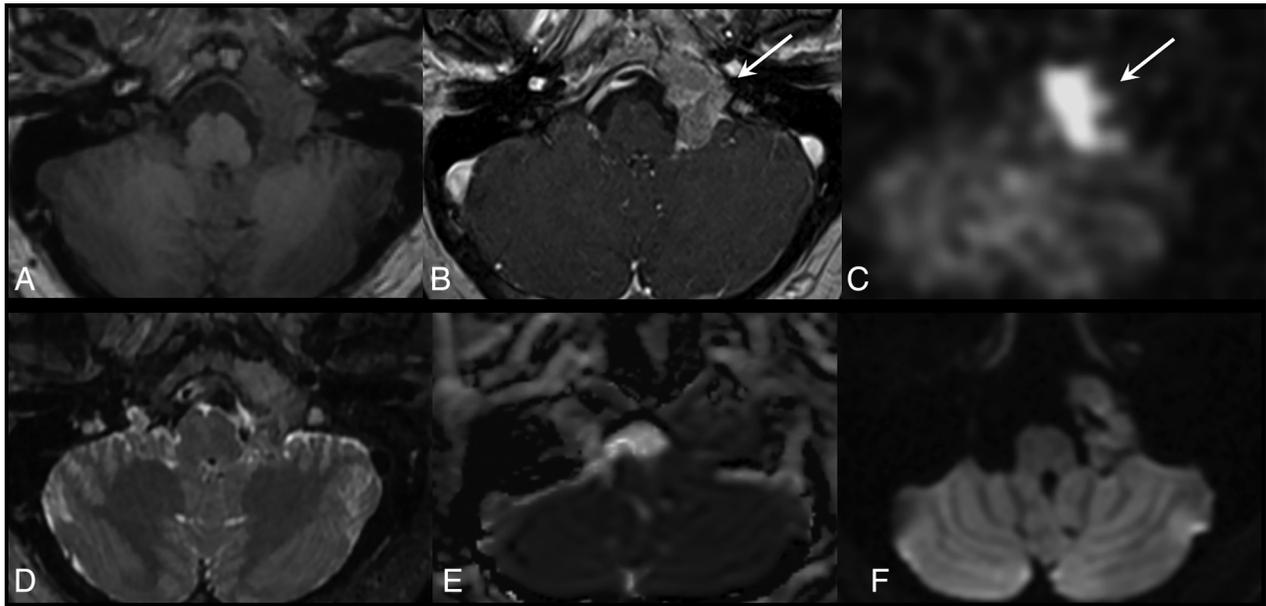
Using the consensus read as the reference, the clinical radiology report at the time of the MR examination demonstrated 100% (30/30; 95% CI, 100%–100%) sensitivity and 93% (50/54;

95% CI, 89%–100%) specificity for identification of paragangliomas (Table 3).

## DISCUSSION

This retrospective study demonstrated that asymmetrically elevated ASL blood flow in the skull base or neck had a sensitivity of 97%–100% and specificity of 74% in detecting HNPGs in this selected cohort. Asymmetrically elevated ASL blood flow also demonstrated almost perfect interrater reliability between 2 neuroradiologists with 12 and 5 years of experience. This study provides preliminary data to support the use of ASL to detect HNPGs.

Most HNPGs are parasympathetic, and <4% may demonstrate sympathetic function, secreting catecholamines.<sup>1,14</sup> Symptoms of a nonfunctional HNPG mostly depend on its anatomic locations, given its slow growth.<sup>15</sup> Carotid body paragangliomas, located at carotid bifurcations, are often asymptomatic until they become large masses, inducing cranial nerve dysfunction or presenting as a neck mass. Although they have a low malignant potential,<sup>16</sup> they are preferably treated with surgical resection. Vagal paragangliomas located along the course of the vagal nerve may present with various symptoms, including pulsatile tinnitus, cranial nerve deficits such as hoarseness (X), dysphagia (IX), shoulder drop (XI), aspiration, and hemiatrophy of the tongue (XII). These tumors may also be asymptomatic. Radiosurgical treatment is often



**FIG 3.** Pathology-proved nonparaganglioma with associated ASL blood flow elevation. A male patient in his 40s initially presented with a headache. Preoperative T1WI precontrast (A) and T1WI postcontrast (B) MR shows an enhancing mass with osseous remodeling. C, ASL blood flow map shows associated blood flow elevation (white arrows). D, T2WI shows that the mass demonstrates signal isointensity to cerebellar gray matter. ADC (E) and DWI (F) demonstrate mild reduced diffusion. The initial clinical diagnosis was paraganglioma, while surgical pathology revealed meningioma.

preferred, given the high risk for vagal nerve injury, with an open approach. Tympanic paragangliomas located in the middle ear along the Jacobson nerve induce pulsatile tinnitus and hearing loss. Jugular paragangliomas, centered at the jugular foramen in the skull base, can occlude venous outflow and induce pulsatile tinnitus, hearing loss, and dizziness. Jugular paragangliomas can also cause cranial nerve deficits when they grow or invade the adjacent cranial nerves.

While HNPGs can arise sporadically, a considerable portion of them are associated with genetic syndromes and high-risk mutations,<sup>14</sup> which accounted for 53% of our cohort. Accurate and timely diagnosis of paragangliomas would allow appropriate management, including genetic testing, functional imaging or biopsy, and surgical or radiation treatment, especially in symptomatic cases or cases with tumor growth.<sup>1,17</sup> For asymptomatic or indolent HNPGs in high-risk locations such as the skull base,<sup>1</sup> the diagnosis is often achieved with imaging features alone and tumors managed with active surveillance and radiation therapy, because surgical risks may outweigh the benefit. Therefore, only a limited number of the patients in our cohort had pathologic data to support their HNPG diagnosis.

Contrast-enhanced MR angiograms may add diagnostic value to conventional MR sequences in identifying HNPG, with a reported sensitivity of 100% and specificity of 94% in a small cohort study of 27 patients.<sup>7</sup> MR perfusion can assess blood flow and may aid in differentiating paragangliomas from nonparaganglioma tumors.<sup>1,18</sup> Among common skull base lesions, paragangliomas demonstrate significantly higher ASL perfusion compared with other hypervascular lesions such as meningiomas and metastases, except for hemangioblastomas, which demonstrate similar hyperperfusion.<sup>6</sup>

Multicenter prospective studies with a cohort of 238 cases have shown that conventional MR and contrast-enhanced MR

angiograms have a sensitivity of 90%–95% and specificity of 92%–99% in diagnosing HNPGs.<sup>19</sup> However, the MR angiogram interpretation of HN lesions may be more complicated than perfusion imaging, with a reported interrater reliability of 0.67–0.77.<sup>7</sup> In comparison, our single-center data showed an almost perfect interrater reliability of 0.90.

In our cohort, the clinical radiology report of conventional MR plus ASL imaging at the time of the study had a sensitivity of 100% and specificity of 93% in diagnosing HNPGs, which is like the reported sensitivity and specificity of contrast-enhanced MR angiography, though direct comparison is not allowed, given different cohorts and institutions. With ASL blood flow alone, the sensitivity of detecting paraganglioma is high (97%–100%), indicating that ASL is very useful in excluding the diagnostic possibility of HNPGs. However, ASL blood flow elevation is not particularly specific in diagnosing paraganglioma, because other common skull base lesions such as meningiomas, hemangiomas, and osseous metastases may also demonstrate hyperperfusion.<sup>6</sup> Therefore, despite its high sensitivity for HNPGs, we do not recommend using ASL alone for the diagnosis of HN tumors. Instead, its value lies in helping increase the sensitivity and diagnostic confidence of radiologists as they synthesize all available conventional imaging and clinical information in the assessment of potential HNPGs.

The study has several limitations. First, some of the cases had no pathology or dotatate PET confirmation of paraganglioma diagnosis due to various practical reasons, such as small lesion size and preference for conservative management. However, each case was analyzed using all available imaging and clinical information and follow-up data to maximize diagnostic accuracy. Second, the examinations were performed on a variety of MR scanners from different manufacturers. Therefore, parameters for ASL and other sequence acquisitions were not unified across all

examinations. However, this variability reflects the real-world practice of most large hospital systems that use multiple scanner systems. The high interrater reliability and sensitivity achieved despite this variability speaks to the value of the ASL technique in detecting HNPGs. Third, although we have blinded the readers to the clinical and final diagnosis of each case, they were nonetheless aware of the selective nature of the study and the study cohort, which was enriched for patients with HNPGs. Therefore, they may have been subconsciously biased toward diagnosing HNPGs, and more real-world validation in the future would be valuable to conclusively establish the high sensitivity found in this study. Fourth, the ASL sequence was qualitatively evaluated in this study. A quantitative method may provide more diagnostic value and may be investigated in future studies.

Finally, this was a retrospective study targeting MR studies that were performed for the evaluation of suspected paragangliomas or that had patients with a diagnosis of paragangliomas, enriching the data set with a higher prevalence of paragangliomas than in the general population. This highly selective cohort did not include all patients with head and neck masses that would demonstrate elevated perfusion; as a result, the reported specificity value of ASL asymmetry in identifying HNPGs may not apply to a more generalized, global cohort. Therefore, caution should be used when interpreting the specificity in this study. In addition, this study cannot provide positive predictive values or negative predictive values of ASL in the evaluation of paragangliomas to reflect real-world data. However, given that asymmetric elevation of ASL blood flow had a near-100% sensitivity for paragangliomas, the presence of this finding should raise a high suspicion for an interpreting radiologist even if no abnormality was initially identifiable on conventional imaging.

## CONCLUSIONS

We demonstrated that asymmetrically elevated ASL blood flow has a high sensitivity and almost perfect interrater reliability in identifying HNPGs in a single-center cohort with suspected or diagnosed paragangliomas. This study provides preliminary data to support the use of ASL to detect HNPGs. However, ASL blood flow elevation was not particularly specific for HNPGs, and conventional imaging and clinical data should be considered when making a final diagnosis. In the future, larger studies are warranted for further validation.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Lin EP, Chin BB, Fishbein L, et al. **Head and neck paragangliomas: an update on the molecular classification, state-of-the-art imaging, and management recommendations.** *Radiol Imaging Cancer* 2022;4:e210088 CrossRef Medline
2. Pellitteri PK, Rinaldo A, Myssiorek D, et al. **Paragangliomas of the head and neck.** *Oral Oncol* 2004;40:563–75 CrossRef Medline
3. Chang CA, Pattison DA, Tothill RW, et al. **(68)Ga-DOTATATE and (18)F-FDG PET/CT in paraganglioma and pheochromocytoma: utility, patterns and heterogeneity.** *Cancer Imaging* 2016;16:22 CrossRef Medline
4. Rini JN, Keir G, Caravella C, et al. **Somatostatin receptor-PET/CT/MRI of head and neck neuroendocrine tumors.** *AJNR Am J Neuroradiol* 2023;44:959–66 CrossRef Medline
5. Amar L, Pacak K, Steichen O, et al. **International consensus on initial screening and follow-up of asymptomatic SDHx mutation carriers.** *Nat Rev Endocrinol* 2021;17:435–44 CrossRef Medline
6. Geerts B, Leclercq D, Tezenas Du Montcel S, et al. **Characterization of skull base lesions using pseudo-continuous arterial spin labeling.** *Clin Neuroradiol* 2019;29:75–86 CrossRef Medline
7. Neves F, Huwart L, Jourdan G, et al. **Head and neck paragangliomas: value of contrast-enhanced 3D MR angiography.** *AJNR Am J Neuroradiol* 2008;29:883–89 CrossRef Medline
8. Ota Y, Liao E, Capizzano AA, et al. **Diagnostic role of diffusion-weighted and dynamic contrast-enhanced perfusion MR imaging in paragangliomas and schwannomas in the head and neck.** *AJNR Am J Neuroradiol* 2021;42:1839–46 CrossRef Medline
9. Wong EC. **An introduction to ASL labeling techniques.** *J Magn Reson Imaging* 2014;40:1–10 CrossRef Medline
10. Haller S, Zaharchuk G, Thomas DL, et al. **Arterial spin-labeling perfusion of the brain: emerging clinical applications.** *Radiology* 2016;281:337–56 CrossRef Medline
11. Puig O, Henriksen OM, Vestergaard MB, et al. **Comparison of simultaneous arterial spin labeling MRI and (15)O-H(2)O PET measurements of regional cerebral blood flow in rest and altered perfusion states.** *J Cereb Blood Flow Metab* 2020;40:1621–33 CrossRef Medline
12. Jarnum H, Steffensen EG, Knutsson L, et al. **Perfusion MRI of brain tumours: a comparative study of pseudo-continuous arterial spin labelling and dynamic susceptibility contrast imaging.** *Neuroradiology* 2010;52:307–17 CrossRef Medline
13. McHugh ML. **Interrater reliability: the kappa statistic.** *Biochem Med (Zagreb)* 2012;22:276–82 CrossRef
14. Richter S, Qiu B, Ghering M, et al. **Head/neck paragangliomas: focus on tumor location, mutational status and plasma methoxytyramine.** *Endocr Relat Cancer* 2022;29:213–24 CrossRef Medline
15. Offergeld C, Brase C, Yaremchuk S, et al. **Head and neck paragangliomas: clinical and molecular genetic classification.** *Clinics (Sao Paulo)* 2012;67(Suppl 1):19–28 CrossRef Medline
16. Taieb D, Kaliski A, Boedeker CC, et al. **Current approaches and recent developments in the management of head and neck paragangliomas.** *Endocr Rev* 2014;35:795–819 CrossRef Medline
17. Graham NJ, Smith JD, Else T, et al. **Paragangliomas of the head and neck: a contemporary review.** *Endocr Oncol* 2022;2:R153–62 CrossRef Medline
18. Gaddikeri S, Hippe DS, Anzai Y. **Dynamic contrast-enhanced MRI in the evaluation of carotid space paraganglioma versus schwannoma.** *J Neuroimaging* 2016;26:618–25 CrossRef Medline
19. Gimenez-Roqueplo AP, Caumont-Prim A, Houzard C, et al. **Imaging work-up for screening of paraganglioma and pheochromocytoma in SDHx mutation carriers: a multicenter prospective study from the PGL-EVA Investigators.** *J Clin Endocrinol Metab* 2013;98:E162–73 CrossRef Medline

# Perfusion Showdown: Comparison of Multiple MRI Perfusion Techniques in the Grading of Pediatric Brain Tumors

 Chang Y. Ho, Nucharin Supakul, Greg Anthony,  Qiuting Wen,  Paula Chen,  Maryann Allen, Jason G. Parker, Benjamin R. Gray, Eric Chen, Paul R. Territo, Scott A. Persohn, Stephen F. Kralik, and Gary D. Hutchins



## ABSTRACT

**BACKGROUND AND PURPOSE:** There are multiple MRI perfusion techniques, with limited available literature comparing these techniques in the grading of pediatric brain tumors. For efficiency and limiting scan time, ideally only one MRI perfusion technique can be used in initial imaging. We compared DSC, dynamic contrast enhancement (DCE), and intravoxel incoherent motion (IVIM) along with ADC from DWI for differentiating high- versus low-grade pediatric brain tumors.

**MATERIALS AND METHODS:** Presurgical MRI at a single pediatric hospital for primary brain tumor of glial, neuronal, or embryonal origin including DWI, DSC, DCE, and IVIM with postsurgical histopathology results with World Health Organization (WHO) tumor grading were included. Tumor VOI was drawn on T2-weighted images registered to selected parametric maps from DWI, DSC, DCE, and IVIM. Multiple Bonferroni-corrected *t* tests were performed on the mean tumor values for each selected parameter, including histogram analysis of 95th percentile relative CBV (rCBV), 5th percentile ADC, and 5th percentile D, to evaluate for significance between high- and low-grade tumors. Receiver operating characteristic (ROC) analysis was performed for significant *t* test results.

**RESULTS:** Seventy-two subjects were included with 36 low-grade and 36 high-grade tumors. *t* test after Bonferroni correction yielded significant results for 5th percentile ADC ( $P = .003$ ), mean ADC ( $P = .006$ ), mean D ( $P = .009$ ), and 5th percentile D ( $P = .02$ ). Specifically, 95th percentile rCBV, mean rCBV,  $D^*$ ,  $f$ , and all DCE parameters were not significant for high- versus low-grade pediatric brain tumors after correction. Only mean rCBV was significant before but not after Bonferroni correction ( $P = .04 \rightarrow P = .35$ ). ROC analysis for the parameters with *t* test significance before and after Bonferroni correction had the following AUC, in descending order: 5th percentile ADC (0.791, 0.68–0.88,  $P < .001$ ), 5th percentile D (0.789, 0.68–0.88,  $P < .001$ ), mean D (0.76, 0.64–0.85,  $P < .001$ ), mean ADC (0.754, 0.64–0.85,  $P < .001$ ), and mean rCBV (0.683, 0.56–0.79,  $P = .004$ ).

**CONCLUSIONS:** Perfusion parameters from IVIM, DCE, and DSC were not able to significantly differentiate between high- versus low-grade pediatric brain tumors compared with ADC in our study. Fifth percentile ADC performed the best and can be the primary technique for grading pediatric brain tumors.

**ABBREVIATIONS:** AUC = area under curve; DCE = dynamic contrast enhancement; IVIM = intravoxel incoherent motion; rCBV = relative CBV; ROC = receiver operating characteristic; WHO = World Health Organization

Primary pediatric brain tumors are the most common solid cancer in children and are a leading cause of cancer morbidity and mortality.<sup>1</sup> Tumor histology is heterogeneous, both in the cell of origin, such as embryonal, glial, and/or neuronal, as well as

the grading of these tumors from low- to high-grade. Tumor grading is typically performed at the histologic level by using the World Health Organization (WHO) classification from grades 1 to 4, with the higher the number typically denoting a more aggressive tumor.<sup>2</sup>

Radiologic biomarkers correlating with tumor grade may be helpful for prognosis, guiding biopsy if gross total resection cannot be safely achieved, and for radiopathologic agreement. A discordant radiopathologic result may require further molecular or genetic testing in certain tumors.<sup>1,2</sup>

ADC from DWI quantifies the diffusion of water molecules and is well recognized for predicting high versus low tumor grades in both adult and pediatric CNS tumors. This is thought to be due to a decrease in water diffusion associated with hypercellular portions of tumors, resulting in a decrease in ADC values.<sup>3-5</sup>

Received September 9, 2024; accepted after revision December 18.

From the Departments of Radiology and Imaging Sciences (C.Y.H., N.S., G.A., Q.W., P.C., M.A., J.G.P., B.R.G., P.R.T., G.D.H.), and Department of Medicine (E.C., P.R.T., S.A.P.), Indiana University School of Medicine, Indianapolis, Indiana; Stark Neuroscience Research Institute (P.R.T., S.A.P.), Indiana University School of Medicine, Indianapolis, Indiana; Department of Diagnostic Medicine (C.Y.H.), The University of Texas at Austin Dell Medical School, Austin, Texas; and Department of Radiology (S.F.K.), Texas Children's Hospital, Houston, Texas.

Please address correspondence to Chang Y. Ho, MD, Department of Diagnostic Medicine, Dell Medical School, 1601 Trinity St., Bldg. B, Stop Z0300, Austin, TX 78712; e-mail: chang.ho@austin.utexas.edu



Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8635>

## SUMMARY

**PREVIOUS LITERATURE:** Prior literature is limited on the comparison of different perfusion techniques in grading pediatric brain tumors. Because of the heterogeneity of common pediatric brain tumors compared with adults, prior literature (by using 1 or 2 different techniques) has shown mixed results in the ability of DSC, DCE, or IVIM techniques to significantly differentiate high- versus low-grade tumors.

**KEY FINDINGS:** When directly compared with ADC from DWI, no perfusion parameter from DSC, DCE, or IVIM performed as well as ADC by using ROC analysis.

**KNOWLEDGE ADVANCEMENT:** ADC from DWI performs better than any perfusion parameter in our cohort and suggests that perfusion MRI techniques may not be necessary when the question is only differentiating high- versus low-grade tumors.

MRI perfusion imaging has also been shown to differentiate tumors in terms of high- versus low-grade, theoretically, due to highly aggressive tumors requiring greater blood supply from increased metabolic demand.<sup>6</sup> Contrast-based techniques, such as DSC and dynamic contrast enhancement (DCE), are complementary in that DSC is a first-pass bolus technique with cerebral blood volume as the main parameter best correlating with tumor grade, while DCE is a permeability technique with  $K^{trans}$  as one of the parameters showing differences between low- and high-grade.<sup>7</sup> Noncontrast techniques include arterial spin labeling,<sup>8</sup> and the diffusion-based intravoxel incoherent motion (IVIM), with parameters such as  $D$ ,  $D^*$ , and  $f$ .  $D^*$  and  $f$  are thought to be associated with CBF and CBV, respectively.<sup>9</sup> However, only one, if any, perfusion sequence is ideally necessary as efficiency in MRI scan time is ideal in clinical scenarios regarding patient tolerance, safety, and throughput.

Our aim is to compare DWI with several perfusion techniques, DSC, DCE, and IVIM, to differentiate low- versus high-grade tumors in a single MRI scan for pediatric patients presenting with primary brain tumors.

## MATERIALS AND METHODS

### Subjects

This is an institutional review board-approved retrospective study from August 2016 to May 2020, with a waiver of informed consent. STARD 2015 checklist was followed where applicable (Supplemental Data). Inclusion criteria included 1) consecutive primary brain tumor presenting at a children's hospital, 2) MRI with DWI, DSC, DCE, and IVIM performed before any surgical intervention or chemoradiation, and 3) histopathology results with WHO 2021<sup>2</sup> tumor grading. Exclusion criteria included tumors not of glial, neuronal, or embryonal origin, such as craniopharyngiomas and germ cell tumors, and those outside of the blood-brain barrier, such as choroid plexus and pineal gland tumors. Motion-limited scans were also excluded by the neuroradiologist (C.Y.H.).

### Anatomic Images

All MRI examinations were performed on 3T scanners (Magnetom Skyra, Siemens Healthineers). Anatomic sequences include axial T2 (4-mm slice thickness,  $358 \times 458$  matrix), axial FLAIR, 3D T1 pre- and postcontrast, susceptibility-weighted imaging, and DWI. DWI (4-mm slice thickness,  $1.5 \times 1.5$  mm<sup>2</sup>

in-plane resolution,  $144 \times 144$  matrix) was performed at  $b = 0$  and  $b = 1000$  seconds/mm<sup>2</sup> in 3 orthogonal planes with ADC maps generated by the scanner. IVIM imaging was performed before contrast injection. Two full doses of gadobenate dimeglumine (0.1 mmol/kg) were given with a power injector, followed by a normal saline flush for a total volume of 32 mL. When possible, an 18- or 20-gauge peripheral intravenous access was used with a power injector rate of 5 mL/s. In some cases, primarily with smaller children, only 24-gauge peripheral intravenous access was possible. The DCE sequence was performed after the first contrast bolus, followed by DSC for the second bolus.<sup>10</sup>

### IVIM Sequence and Processing

DWI for IVIM fitting was conducted in axial planes by using 8  $b$ -values (0, 50, 100, 150, 200, 400, 600, and 800 seconds/mm<sup>2</sup>), each acquired at 3 orthogonal diffusion directions with a monopolar diffusion gradient. The DWI sequence parameters were configured as follows: TR/TE: 3400/70 ms, matrix:  $130 \times 130$ , in-plane resolution:  $1.5 \times 1.5$  mm<sup>2</sup>, slice thickness: 4 mm, slices: 27, and acquisition time = 2:28 minutes.

IVIM analysis was carried out by using a stepwise biexponential fitting method through a publicly available toolbox.<sup>11,12</sup> Three maps were generated, including the true diffusion coefficient ( $D$ , mm<sup>2</sup>/s), pseudo-diffusion coefficient ( $D^*$ , mm<sup>2</sup>/s), and a fraction of tissue volume with incoherent perfusion ( $f$ ).

### Dynamic Contrast Enhancement Imaging and Processing

Dynamic contrast-enhanced images were acquired by using a T1-weighted gradient-echo sequence (TR/TE: 1.54/3.91 ms, flip angle: 10°, slice thickness: 5 mm, matrix:  $154 \times 192$ , slices: 20, time points: 100). One hundred time point images ( $S(t)$ ) were acquired over 4.5 minutes. An in-house parametric image software package (Qimage) developed in IDL (Exelis Visual Information Solutions, a subsidiary of Harris Corporation) was utilized to analyze all studies. The first 6 image frames were averaged to generate a noncontrast image ( $S_0$ ). Images proportional to gadolinium concentration were estimated as  $C(t) = S(t) - S_0$ . ROIs were placed over the right MCA on a bolus phase gadolinium concentration image to define the arterial input function. The left MCA was utilized when the right MCA was not large enough for ROI placement or tumor anatomy displaced the right MCA. Arterial input function ROI placement was verified by a pediatric neuro-radiologist (C.Y.H.). Arterial input function concentration curves

were corrected for differences in gadolinium relaxivity between tissue ( $r_1 = 4.5 \text{ seconds}^{-1}$ ) and blood ( $r_1 = 1.7 \text{ seconds}^{-1}$ ). Perfusion parametric images were generated by using a 2-stage fit process to each voxel concentration curve to control for ill-conditioned parameter estimation problems in normal brain tissue. The first stage fits the bolus phase of the study (first 20 seconds following bolus arrival) to estimate the blood volume fraction in each voxel. In the second stage, the blood volume fraction times the arterial input function curve is subtracted from the voxel concentration curve, resulting in a residual concentration curve. Each voxel residual curve is then fit with the expanded Tofts model ( $K^{\text{trans}}$ ,  $V_e$ ,  $K_{ep}$ ,  $V_p$ ) to generate parametric images.<sup>13</sup>

### Dynamic Susceptibility Contrast Imaging and Processing

DSC images were acquired by using a gadolinium contrast-enhanced T2\*-weighted echo-planar image sequence (Siemens ep2d\_perf\_p2; TR/TE: 1600/30 ms, flip angle: 90°, slice thickness: 4 mm, matrix size: 128 × 128, slices: 20). Sixty measurements were acquired at approximately 1.7 seconds per measurement. DSC data sets were postprocessed by using commercially available software (syngo Neuro Perfusion, Siemens Healthineers) to generate CBV maps, which employ leakage correction.<sup>14</sup> Relative CBV (rCBV) was calculated for each subject as the ratio of the CBV in the tumor and normal-appearing white matter ROIs drawn by the neuroradiologist (C.Y.H.).

### Image Registration

Each dynamic image data set (DSC, DCE) was motion-corrected by using the “mcfliirt” command in FSL<sup>15</sup> before applying perfusion modeling or registration. All image series were spatially registered to T2-weighted images by using the “flirt” command in FSL (12° of freedom, -90° to +90° rotation search window, sinc interpolation with 7-pixel width, Hanning window). Correlation ratio was used as the registration cost function unless registration accuracy was inadequate, in which case mutual information was used. When compared with correlation, mutual information maximization enabled better registration of images with similar anatomic structures but different contrast.<sup>16</sup> Registration accuracy was assessed for each image contrast by using the FSL “slicer” command to visualize slices sampled at 35%, 45%, 55%, and 65% of the FOV along each axis. Outlines of boundaries in the T2-weighted reference images were overlaid with registered images (and vice versa) to confirm adequate spatial registration. Parametric maps (ADC,  $K^{\text{trans}}$ ,  $V_e$ ,  $K_{ep}$ ,  $V_p$ , CBV, D,  $D^*$ , and  $f$ ) were registered to T2-weighted images by applying the same transform matrix as the raw image sets.

### Tumor Voxel of Interest

Tumor VOIs were manually drawn by the neuroradiologist (C.H.), a neuroradiology fellow (B.R.G.), or a medical student (E.C.) blinded to pathology on the T2-weighted images, including all solid tumors at each slice, excluding cysts, hematoma, and major vessels, by using the pre- and postcontrast 3D T1 images for reference (Fig 1). All tumor VOIs were checked for accuracy by the neuroradiologist (C.Y.H.). A 2D ROI was also obtained in normal-appearing white matter by the neuroradiologist (C.Y.H.). Normal-appearing white matter was drawn in a single slice of the

cerebral white matter with as large an area as possible, avoiding edema from hydrocephalus and appreciable vessels.

### Image Data Extraction

Numerical image data statistics from each parametric map were extracted by using the “fslstats” command. Tumor VOI-averaged values were extracted from registered data sets by using tumor and normal-appearing white matter ROIs drawn in the T2-weighted reference images. Parametric fitting resulted in some artifactual zero pixels in the ROIs, which were excluded from the averages.

### Data Analysis

Resulting mean tumor values for parameters from DCE ( $K^{\text{trans}}$ ,  $V_e$ ,  $K_{ep}$ ,  $V_p$ ), IVIM (D,  $D^*$ ,  $f$ ), DSC (rCBV), and DWI (ADC) were compared by using 2-tailed Student *t* test (Excel for Microsoft 365, 2022; Microsoft) for high- and low-grade tumor groups based on WHO grading from pathology. In addition to mean tumor values, histogram analysis with 95th percentile rCBV, 5th percentile ADC, and 5th percentile D was also evaluated with Student *t* test. WHO grades I and II were grouped as low-grade, and WHO III and IV were grouped as high-grade. As a result of multiple *t* test comparisons, the Holm-Bonferroni correction was performed. Receiver operating characteristic (ROC) analysis (MedCalc Statistical Software Version 23.0.5, MedCalc Software) with area under curve (AUC) was performed for significant *t* test results before and after correction.

## RESULTS

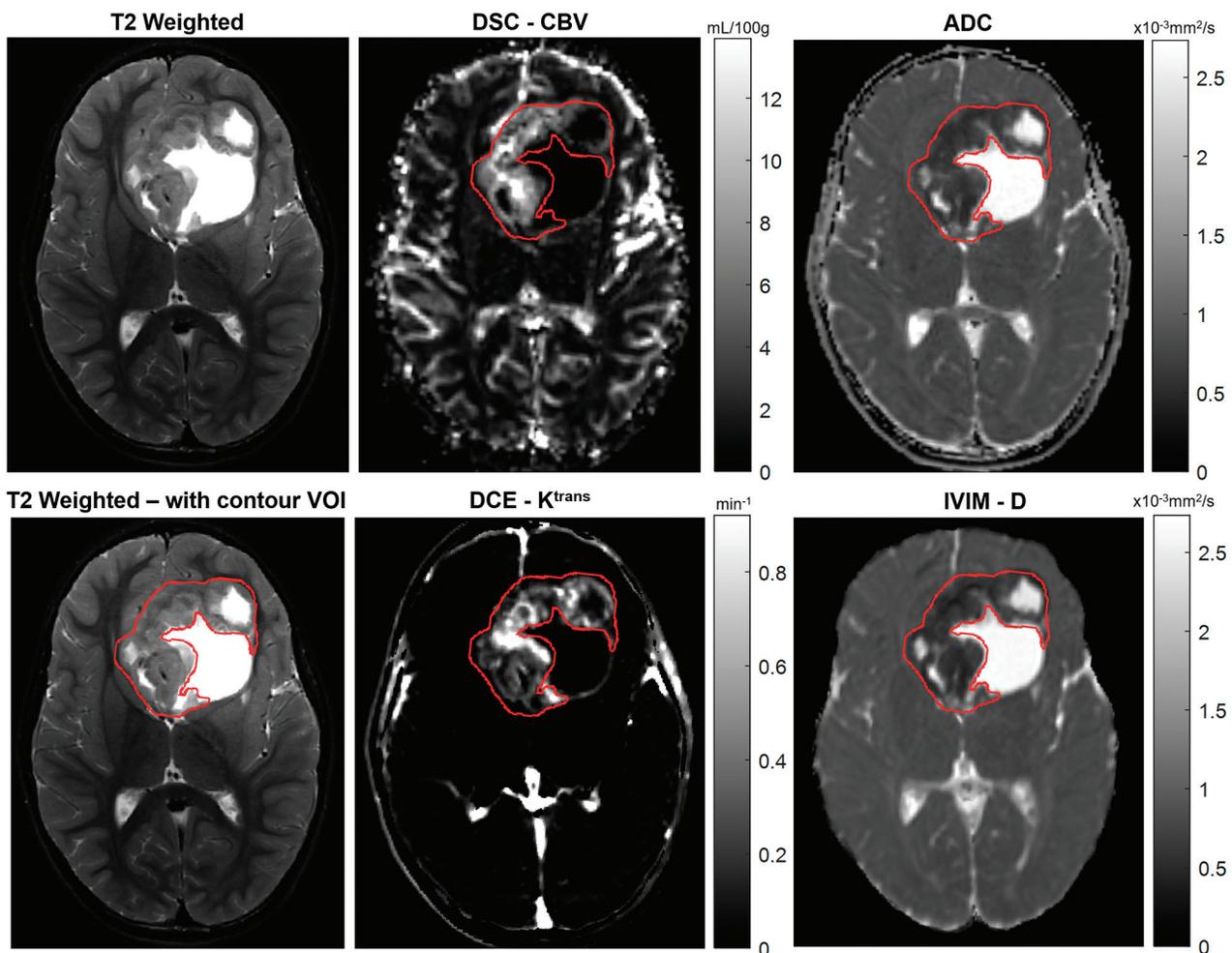
Seventy-two subjects (1.7–211 months, mean 82.8 months, 44 males) met the inclusion and exclusion criteria. There were 36 low-grade and 36 high-grade tumors with tumor histology detailed in Table 1. Twenty-three subjects were excluded because of lack of all perfusion methods ( $n = 9$ ); pathology that was not of glial, neuronal, or embryonal origin ( $n = 11$ ); or excessive motion ( $n = 3$ ).

Student *t* test after Bonferroni correction only yielded significant results for 5th percentile ADC ( $P = .003$ ), mean ADC ( $P = .006$ ), mean D ( $P = .009$ ), and 5th percentile D ( $P = .02$ ). Specifically, 95th percentile rCBV, mean rCBV,  $D^*$ ,  $f$ , and all DCE parameters were not significant for high- versus low-grade pediatric brain tumors after statistical correction. Only mean rCBV was significant before but not after Bonferroni correction ( $P = .04 \rightarrow P = .35$ ) (Table 2).

ROC analysis for the parameters with *t* test significance before and after correction had the following AUC in descending order: 5th percentile ADC (0.791, 0.68–0.88,  $P < .001$ ), 5th percentile D (0.789, 0.68–0.88,  $P < .001$ ), mean D (0.76, 0.64–0.85,  $P < .001$ ), mean ADC (0.754, 0.64–0.85,  $P < .001$ ), and mean rCBV (0.683, 0.56–0.79,  $P = .004$ ). Figures 2–6 demonstrate the sensitivity and specificity of each ROC curve at Youden index.

## DISCUSSION

In this study, we showed that 5th percentile ADC best predicts high- versus low-grade primary pediatric CNS tumors compared with rCBV from DSC or any perfusion-based parameters from DCE or IVIM. In fact, in *t* test analysis of perfusion parameters, only mean rCBV had significance before Bonferroni correction,



**FIG 1.** Subject with supratentorial ependymoma. T2-weighted image, T2 with contour VOI, and parametric maps with registered VOI contours with associated scales.

**Table 1: Number of subjects by histology and WHO grade of included pediatric brain tumors**

Pathology	Number	WHO Grade
Pilocytic astrocytoma	21	1
Medulloblastoma	13	4
Atypical teratoid/rhabdoid tumor	7	4
Anaplastic ependymoma	6	3
Diffuse midline glioma	4	4
Dysembryoplastic neuroepithelial tumor	4	1
CNS embryonal tumor NOS	2	4
Ganglioglioma	2	2
High-grade glioma	2	3
Anaplastic astrocytoma	1	3
Desmoplastic infantile ganglioglioma	1	1
Diffuse astrocytoma	1	2
Ependymoma	1	2
Ganglioblastoma	1	4
Low-grade astrocytoma	1	2
Low-grade glial neoplasm	1	2
Low-grade neuroepithelial neoplasm	1	2
Optic chiasm glioma	1	1
Piloxyoid astrocytoma	1	1
Pleomorphic xanthoastrocytoma	1	2
Total	72	

**Note:**—NOS indicates Not Otherwise Specified.

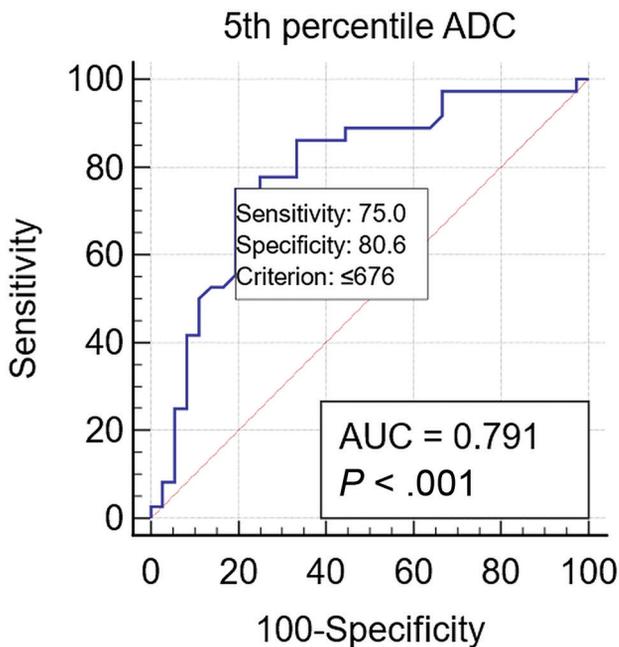
with a loss of significance after correction. In ROC analysis, 5th percentile ADC had the greatest AUC with 5th percentile and mean D from IVIM and mean ADC having similar AUC values. The significance of the D parameter from IVIM is not surprising as D is the closest to clinical ADC, measured between  $b = 0$  and  $b = 1000$  seconds/mm<sup>2</sup>, with D\* and f derived from the curve in the lower b-values and theoretically, microscopic perfusion from the capillary network. In typical clinical workflows, where histogram analysis or complete tumor VOI is impractical, the “low spot” technique of choosing small ROIs of the subjectively lowest ADC value may be a robust method for predicting high versus low tumor grade.<sup>17</sup> Additional IVIM sequences for this purpose would be unnecessary.

The lack of significance in any perfusion-based parameter in this study may be due to the heterogeneity of primary pediatric brain tumors. Compared with adults, where glial and glioneuronal tumors dominate, the most common primary pediatric brain tumors are glial pilocytic astrocytomas and embryonal medulloblastomas.<sup>1</sup> As described by Ho et al<sup>18</sup> in previous papers, there is considerable overlap in the range of rCBV from DSC between the 2 tumors. In fact, supratentorial and pleomorphic variants of pilocytic astrocytomas have significantly increased rCBV from DSC compared with classic infratentorial pilocytic astrocytomas.<sup>19</sup>

**Table 2: t test means and P values for ADC and perfusion parameters<sup>a</sup>**

	Mean Low Grade	Standard Deviation Low Grade	Mean High Grade	Standard Deviation High Grade	t Test P Value	Bonferonni-Corrected P Value
5th % ADC	987.97	403.63	659.47	301.06	<b>.0002</b>	<b>.0026</b>
Mean ADC	1423.44	433.31	1072.31	387.87	<b>.0006</b>	<b>.0061</b>
Mean D	0.0014	0.0004	0.0011	0.0004	<b>.0009</b>	<b>.0090</b>
5th % D	0.0010	0.0004	0.0007	0.0004	<b>.0027</b>	<b>.024</b>
Mean rCBV	1.75	1.27	2.39	1.39	<b>.044</b>	.35
95th % rCBV	4.15	2.92	5.78	4.013	.053	.38
Mean $k_{ep}$	18.28	0.45	0.71	0.52	.075	.45
Mean $K_{trans}$	406.17	13.79	193.02	1013.58	.29	1.0000
Mean $V_p$	2.70	0.058	1.56	8.60	.31	1.0000
Mean $V_e$	31.62	0.13	0.84	0.18	.36	1.0000
Mean f	0.063	0.042	0.057	0.026	.50	1.0000
Mean D*	0.022	0.007	0.022	0.0058	.72	1.0000

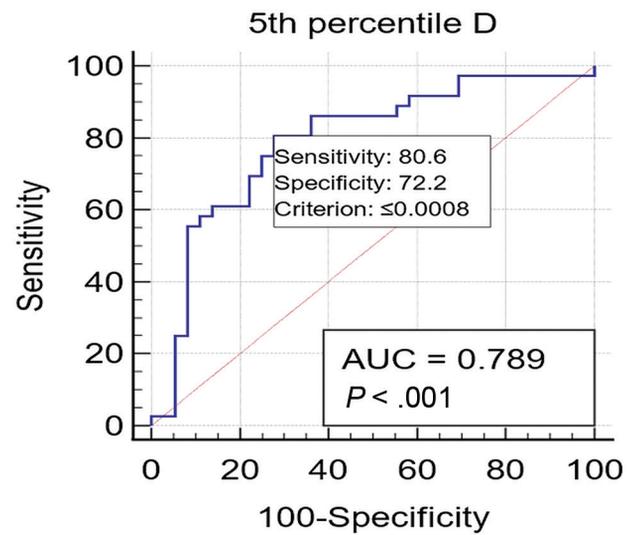
<sup>a</sup>Bolded values are significant ( $P < .05$ ).



**FIG 2.** ROC curve for 5th percentile ADC with Youden index ( $10^{-6} \text{ mm}^2/\text{s}$ ).

Histologically, pilocytic astrocytomas can also have high vascularity with hyalinized and glomeruloid vessels, elevating rCBV,<sup>20,21</sup> which is a confounder in using CBV as a parameter to differentiate high- versus low-grade pediatric brain tumors.

Our results contrast with previous literature in perfusion on grading pediatric brain tumors.<sup>22</sup> In a multicenter study, Withey et al<sup>23</sup> found that rCBV from DSC was significantly different between 40 low-grade and 45 high-grade tumors. This study was limited in that the 7 centers had a nonuniform DSC protocol, including different scanners, magnet strength, and contrast agents, and some centers had a prebolus contrast dose for extracellular space saturation while others did not. Protocol nonuniformity is a limitation as TR, TE, and flip angle can affect T1 and T2 weighting, ultimately affecting CBV measurements.<sup>24</sup> An interesting finding in their study was that software leakage correction and prebolus contrast both increased the rCBV for pilocytic astrocytomas compared with uncorrected rCBV. Our study had both prebolus contrast and leakage correction,

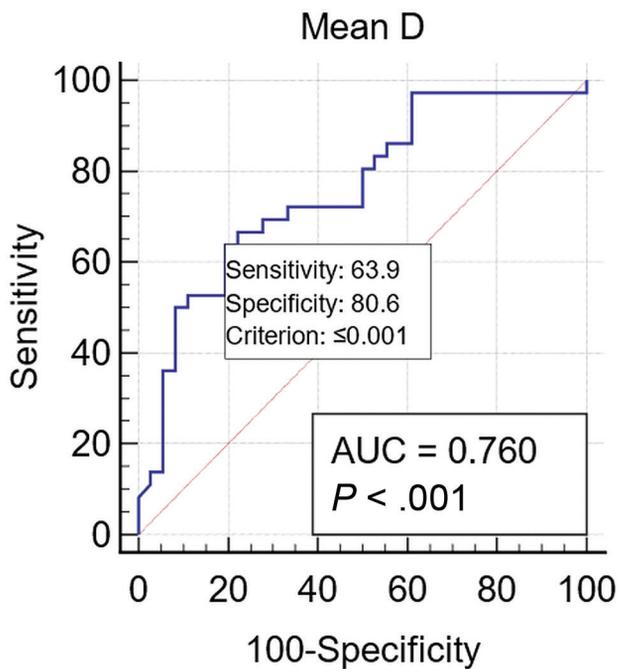


**FIG 3.** ROC curve for 5th percentile D with Youden index ( $\text{mm}^2/\text{s}$ ).

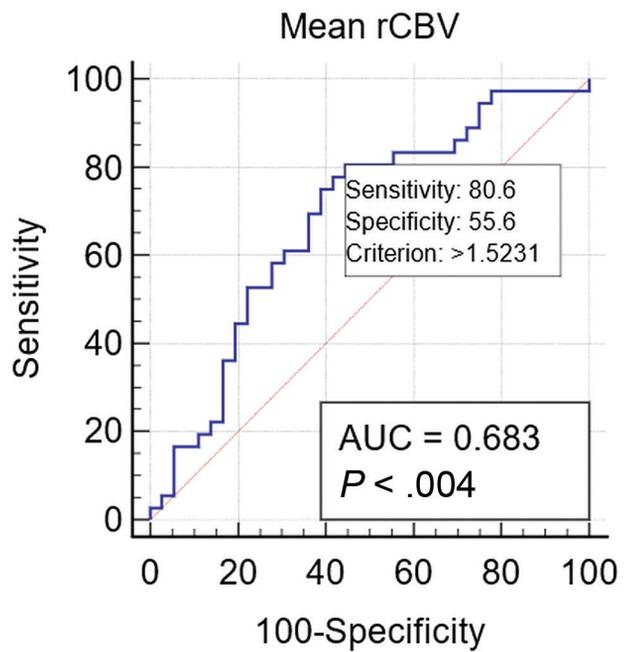
which may contribute to nonsignificance between the low- and high-grade groups.

For DCE, Vajapeyam et al<sup>25</sup> found significance by using linear regression models from low- to high-grade for the  $K_{trans}$ ,  $K_{ep}$ , and  $V_e$  parameters of DCE in 38 patients. It should be noted that linear regression shows the relationship between 2 variables and not whether 2 sample populations are significantly different by the means of the variable. Furthermore, Vajapeyam et al<sup>25</sup> also excluded nonenhancing tumors, which is not uncommon for all tumor types in children. Gupta et al<sup>26</sup> evaluated DCE on 64 subjects but found no significant differences except for  $V_p$  and rCBV. rCBV was not from the typical T2\* technique as in our study but derived from the T1-based DCE technique. Also interesting to note is that in their study of 64 subjects, only 5 were pilocytic astrocytomas, the most common low-grade pediatric brain tumor. In their study, ependymomas and pilocytic astrocytomas had significantly higher rCBV than medulloblastoma, which is contradictory to the theory of increasing cerebral blood volume with tumor grade. This is consistent with the findings of Ho et al.<sup>18</sup>

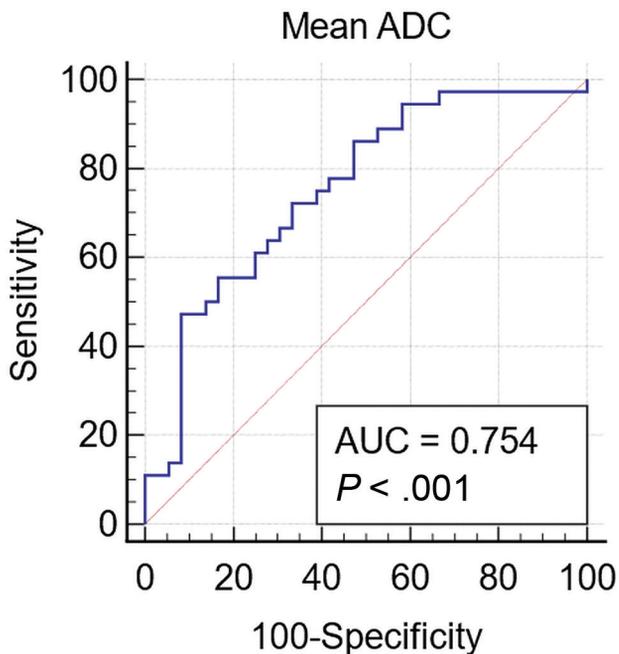
In 1 of 2 other known studies evaluating IVIM in pediatric brain tumors, Kikuchi et al<sup>27</sup> found significant differences



**FIG 4.** ROC curve for mean D with Youden index ( $10^{-6} \text{ mm}^2/\text{s}$ ).



**FIG 6.** ROC curve for mean rCBV with Youden index.



**FIG 5.** ROC curve for mean ADC with Youden index ( $\text{mm}^2/\text{s}$ ).

between 10 low-grade and 7 high-grade tumors for ADC, D, and  $f$  but not  $D^*$ . Our results agree with the significant differences in ADC and D between high- and low-grade tumors.  $D^*$  is theoretically analogous to cerebral blood flow, and  $f$  may reflect tumor vascularity similar to CBV. A smaller number of subjects limits the study by Kikuchi et al,<sup>27</sup> which may include pilocytic astrocytomas on the lower vascularity spectrum. Similar to our study, She et al<sup>28</sup> found no significant difference between high- and low-grade groups in 54 subjects for  $f$  and

$D^*$  and only found a significant difference in ADC, D, and other kurtosis measures.

ASL was not evaluated in this study, though previous studies have demonstrated significance in the normalized CBF derived from ASL between high- and low-grade groups.<sup>8,29,30</sup> Testud et al<sup>31</sup> compared CBF from ASL, CBV from DSC, and ADC by using ROC curves and found that CBF from ASL was not significant and had lower AUC compared with CBV and ADC, with ADC having the highest AUC. This is also consistent with our study in that ADC has superior performance compared with any perfusion parameter from DSC, DCE, or IVIM.

From our study, ADC is robust and easily available in routine clinical sequences and outperforms any perfusion parameter when grading pediatric brain tumors between high and low grade. Though some studies show a benefit in combining ADC with a perfusion metric,<sup>32</sup> the improvement in performance may not be beneficial for clinical throughput and efficiency from a scan time and radiology reading perspective. Furthermore, multiple technical variables can have an impact on perfusion parameters. These include scanning parameters. With regard to DSC, prebolus contrast, T1 effect, and leakage correction may increase the CBV of pilocytic astrocytomas. This makes perfusion comparisons limited across different scanners and institutions without harmonizing techniques.

Rather than grading, perfusion imaging in pediatric brain tumors may be helpful as a biomarker correlating with genetic differences within specific histologic tumor diagnoses, including tumor subtype and ultimate outcome. The large range of both pilocytic and medulloblastoma CBV suggests that this and related perfusion parameters may be useful in correlating with the presence of *BRAF* mutations in pilocytic and the multiple molecular subtypes of medulloblastomas. A larger number of specific tumors and molecular subtypes with uniform technical parameters will be necessary to elucidate these potential differences.

## CONCLUSIONS

Perfusion parameters from IVIM, DCE, DSC were not able to significantly differentiate between high- versus low-grade pediatric brain tumors when compared with ADC in our study. Fifth percentile ADC performed the best and should be the primary technique for grading pediatric brain tumors.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Jaimes C, Poussaint TY. **Primary neoplasms of the pediatric brain.** *Radiology Clin North Am* 2019;57:1163–75 CrossRef Medline
2. Louis DN, Perry A, Wesseling P, et al. **The 2021 WHO Classification of Tumors of the Central Nervous System: a summary.** *Neuro Oncol* 2021;23:1231–51 CrossRef Medline
3. Jaremko JL, Jans LBO, Coleman LT, et al. **Value and limitations of diffusion-weighted imaging in grading and diagnosis of pediatric posterior fossa tumors.** *AJNR Am J Neuroradiol* 2010;31:1613–16 CrossRef Medline
4. Kralik SF, Taha A, Kamer AP, et al. **Diffusion imaging for tumor grading of supratentorial brain tumors in the first year of life.** *AJNR Am J Neuroradiol* 2014;35:815–23 CrossRef Medline
5. Seo M, Choi Y, Soo Lee Y, et al. **Glioma grading using multi-parametric MRI: head-to-head comparison among dynamic susceptibility contrast, dynamic contrast-enhancement, diffusion-weighted images, and MR spectroscopy.** *Eur J Radiology* 2023;165:110888 CrossRef Medline
6. Essig M, Nguyen TB, Shiroishi MS, et al. **Perfusion MRI: the five most frequently asked clinical questions.** *AJR Am J Roentgenol* 2013;201:W495–510 CrossRef Medline
7. Ibrahim M, Ghazi TU, Bapuraj JR, et al. **Contrast pediatric brain perfusion: dynamic susceptibility contrast and dynamic contrast-enhanced MR imaging.** *Magn Reson Imaging Clin N Am* 2021; 29:515–26 CrossRef Medline
8. Yeom KW, Mitchell LA, Lober RM, et al. **Arterial spin-labeled perfusion of pediatric brain tumors.** *AJNR Am J Neuroradiol* 2014; 35:395–401 CrossRef Medline
9. Luo H, He L, Cheng W, et al. **The diagnostic value of intravoxel incoherent motion imaging in differentiating high-grade from low-grade gliomas: a systematic review and meta-analysis.** *Br J Radiology* 2021;94:20201321 CrossRef Medline
10. Morabito R, Alafaci C, Pergolizzi S, et al. **DCE and DSC perfusion MRI diagnostic accuracy in the follow-up of primary and metastatic intra-axial brain tumors treated by radiosurgery with cyberknife.** *Radiat Oncol* 2019;14:65 CrossRef Medline
11. Jalnefjord O, Andersson M, Montelius M, et al. **Comparison of methods for estimation of the intravoxel incoherent motion (IVIM) diffusion coefficient (D) and perfusion fraction (f).** *MAGMA* 2018; 31:715–23 CrossRef Medline
12. Gustafsson O, Montelius M, Starck G, et al. **Impact of prior distributions and central tendency measures on Bayesian intravoxel incoherent motion model fitting.** *Magn Reson Med* 2018;79:1674–83 CrossRef Medline
13. Tofts PS, Parker GJ. **DCE-MRI: acquisition and analysis techniques.** In: Barker PB, Golay X, Zaharchuk G, eds. *Clinical perfusion MRI: techniques and applications.* Cambridge University Press; 2013:58–74 CrossRef
14. Boxerman JL, Schmainda KM, Weisskoff RM. **Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not.** *AJNR Am J Neuroradiol* 2006;27:859–67 Medline
15. Jenkinson M, Beckmann CF, Behrens TE, et al. **FSL.** *Neuroimage* 2012;62:782–90 CrossRef Medline
16. Maes F, Collignon A, Vandermeulen D, et al. **Multimodality image registration by maximization of mutual information.** *IEEE Trans Med Imaging* 1997;16:187–98 CrossRef Medline
17. Darbar A, Waqas M, Enam SF, et al. **Use of preoperative apparent diffusion coefficients to predict brain tumor grade.** *Cureus* 2018;10: e2284 CrossRef Medline
18. Ho CY, Cardinal JS, Kamer AP, et al. **Relative cerebral blood volume from dynamic susceptibility contrast perfusion in the grading of pediatric primary brain tumors.** *Neuroradiology* 2015;57:299–306 CrossRef Medline
19. Ho CY, Supakul N, Patel PU, et al. **Differentiation of pilocytic and pilomyxoid astrocytomas using dynamic susceptibility contrast perfusion and diffusion weighted imaging.** *Neuroradiology* 2020; 62:81–88 CrossRef Medline
20. Sie M, de Bont ESJM, Scherpen FJG, et al. **Tumour vasculature and angiogenic profile of paediatric pilocytic astrocytoma; is it much different from glioblastoma?** *Neuropathol Appl Neurobiol* 2010; 36:636–47 CrossRef Medline
21. Ho CY, Cardinal JS, Kamer AP, et al. **Contrast leakage patterns from dynamic susceptibility contrast perfusion MRI in the grading of primary pediatric brain tumors.** *AJNR Am J Neuroradiol* 2016; 37:544–51 CrossRef Medline
22. Dallery F, Bouzerar R, Michel D, et al. **Perfusion magnetic resonance imaging in pediatric brain tumors.** *Neuroradiology* 2017;59:1143–53 CrossRef Medline
23. Withey SB, MacPherson L, Oates A, et al. **Dynamic susceptibility-contrast magnetic resonance imaging with contrast agent leakage correction aids in predicting grade in pediatric brain tumours: a multicenter study.** *Pediatr Radiology* 2022;52:1134–49 CrossRef Medline
24. Welker K, Boxerman J, Kalnin A; American Society of Functional Neuroradiology MR Perfusion Standards and Practice Subcommittee of the ASFNR Clinical Practice Committee, et al. **ASFNR recommendations for clinical performance of MR dynamic susceptibility contrast perfusion imaging of the brain.** *AJNR Am J Neuroradiol* 2015;36:E41–51 CrossRef Medline
25. Vajapeyam S, Stamoulis C, Ricci K, et al. **Automated processing of dynamic contrast-enhanced MRI: correlation of advanced pharmacokinetic metrics with tumor grade in pediatric brain tumors.** *AJNR Am J Neuroradiol* 2017;38:170–75 CrossRef Medline
26. Gupta PK, Saini J, Sahoo P, et al. **Role of dynamic contrast-enhanced perfusion magnetic resonance imaging in grading of pediatric brain tumors on 3T.** *Pediatr Neurosurg* 2017;52:298–305 CrossRef Medline
27. Kikuchi K, Hiwatashi A, Togao O, et al. **Intravoxel incoherent motion MR imaging of pediatric intracranial tumors: correlation with histology and diagnostic utility.** *AJNR Am J Neuroradiol* 2019; 40:878–84 CrossRef Medline
28. She D, Lin S, Guo W, et al. **Grading of pediatric intracranial tumors: are intravoxel incoherent motion and diffusional kurtosis imaging superior to conventional DWI?** *AJNR Am J Neuroradiol* 2021;42:2046–53 CrossRef Medline
29. Morana G, Tortora D, Stagliano S, et al. **Pediatric astrocytic tumor grading: comparison between arterial spin labeling and dynamic susceptibility contrast MRI perfusion.** *Neuroradiology* 2018;60:437–46 CrossRef Medline
30. Dangouloff-Ros V, Deroulers C, Foissac F, et al. **Arterial spin labeling to predict brain tumor grading in children: correlations between histopathologic vascular density and perfusion MR imaging.** *Radiology* 2016;281:553–66 CrossRef Medline
31. Testud B, Brun G, Varoquaux A, et al. **Perfusion-weighted techniques in MRI grading of pediatric cerebral tumors: efficiency of dynamic susceptibility contrast and arterial spin labeling.** *Neuroradiology* 2021; 63:1353–66 CrossRef Medline
32. Su Y, Kang J, Lin X, et al. **Whole-tumor histogram analysis of diffusion and perfusion metrics for noninvasive pediatric glioma grading.** *Neuroradiology* 2023;65:1063–71 CrossRef Medline

# Comparing CT-Like Bone Images Based on Fast Field Echo Resembling a CT Using Restricted Echo Spacing (FRACTURE) MR with CT in Pediatric Congenital Vertebral Anomalies

Hirva Manek and  Foram Gala

## ABSTRACT

**BACKGROUND AND PURPOSE:** Congenital vertebral anomalies are commonly associated with underlying spinal cord anomaly that necessitates imaging both the spinal cord and the bony vertebral column to understand the extent of the deformity better. While MRI is the standard for spinal cord imaging, it does not provide CT-like bone details. Many MR bone imaging techniques have been tested in various adult spine conditions in the past decade but not much has been described on their reliability in pediatric spine. We elaborate on our experience with fast field echo resembling a CT by using restricted echo spacing (FRACTURE) MR bone imaging in congenital vertebral anomalies in children.

**MATERIALS AND METHODS:** Eleven pediatric patients referred to the imaging department for CT and MR study of congenital vertebral anomaly were prospectively included. After receiving informed consent from these patients' guardians, both studies were performed in a single setting and under a single sedation. FRACTURE MR was accelerated by using the compressed sensitivity encoding (CS) technique to reduce the imaging time. We then compared FRACTURE MR and CT images for image quality and studied parameters such as formation or segmentation anomalies, anomalous shape of vertebrae, and alignment deformities.

**RESULTS:** FRACTURE MR showed acceptable image quality with diagnostically limiting artifacts in only 1 patient. The interreader agreement was perfect in the assessment of vertebral body segmentation or formation anomaly and alignment abnormalities, and it was substantial for posterior element anomalies. The bone signal was lower in children under the age of 3 years due to a more immature and cartilaginous skeleton.

**CONCLUSIONS:** FRACTURE MR provides images of acceptable quality in pediatric spinal anomalies. The addition of this novel sequence can be complementary to conventional MR in providing osseous details and CT can be reserved for certain specific indications like postoperative cases. This can help in reducing the radiation dose to this group of pediatric patients who will be serially followed up with imaging during their management.

**ABBREVIATIONS:** CS = compressed sensitivity encoding; FRACTURE = fast field echo resembling a CT using restricted echo spacing; KF = Klippel Feil; UTE = ultrashort time to echo; ZTE = zero time to echo

Congenital anomalies of vertebral segmentation or fusion are frequently associated with underlying cord abnormalities, ranging from 35%–53%.<sup>1,2</sup> This necessitates imaging both the osseous and neural components of the spinal deformity. MRI is the standard for imaging the spine, provides excellent soft tissue resolution and multiplanar images, and is free of ionizing radiation. However, one major drawback of the conventional MR sequences is the inability to provide bone details because they

lack free water. CT is the current standard for bone imaging, but it uses ionizing radiation. Imaging a patient with 2 different modalities for the comprehensive study of pathology results in complexity in the workflow and patient transfer with increased time and cost. All of these factors together become more important when imaging young children and infants where sedation comes into play to reduce motion artifacts and obtain optimal image quality. This can be streamlined if we can obtain CT-like bone images on MR, but this is a challenge because bone lacks free protons and the dark signal of the cortical bone obtained on conventional MR bone sequences is not specific to it. Routine MR protocols for musculoskeletal imaging employ the use of T1-weighted spine echo (SE) or GRE images for structural bone imaging with fat suppression or water excitation techniques. In recent years, clinical radiology has been introduced to a few MR-

Received September 2, 2024; accepted after revision December 14.

From the Department of Radiology, Bai Jerbai Wadia Hospital for Children, Mumbai, Maharashtra, India.

Please address correspondence to Hirva Manek, St. Michael's Hospital, 36 Queen St E, Toronto, ON M5B 1W8; e-mail: hirvamanek@gmail.com; @HirvaManek; @drforamgala; @wadiahospitals



Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8639>

## SUMMARY

**PREVIOUS LITERATURE:** The feasibility of different MR bone imaging techniques such as T1 Gradient Echo, ultrashort time to echo, zero time to echo, and SWI has been described in the past decade mainly focusing on adult vertebral pathologies. Their results showed acceptable image quality and reliability in morphometric assessments of vertebral disorders. Not much has been described about their use and challenges in pediatric spine, the age group that can benefit the most from avoiding radiation-based imaging techniques. We describe our experience with one such MR bone imaging technique called fast field echo resembling a CT using restricted echo spacing (FRACTURE) MR in pediatric congenital vertebral anomalies and compare it with CT.

**KEY FINDINGS:** The image quality of FRACTURE MR images was acceptable with excellent performance in the assessment of the extent of deformity when compared with standard CT. The bone signal and cortical trabecular differentiation are excellent in older children, but the bone signal is lower in younger children (<3 years) with more cartilaginous bones.

**KNOWLEDGE ADVANCEMENT:** FRACTURE MR can complement conventional MR in comprehensive imaging of vertebral anomalies in children by providing osseous details, reducing the need for additional study, and avoiding radiation. The age of a child can have an impact on the bone signal, but it is not usually limiting in the diagnosis of formation and segmentation anomaly.

based bone imaging techniques that increase bone specificity on MR by providing uniform soft tissue contrast or giving CT-like images. These newer techniques include SWI, ultrashort time to echo (UTE), and zero time to echo (ZTE) techniques, 3D GRE techniques, and synthetic CTs with advanced postprocessing techniques to generate CT-like images.<sup>3</sup> Each of these techniques has its strengths and weaknesses when factors such as large-scale applicability, cost, scanning time and postprocessing time, susceptibility to artifacts, and its ability to differentiate cortical and cancellous bone are considered. The feasibility of these MR bone imaging techniques has been studied in a few adult and pediatric conditions such as head trauma, musculoskeletal pathologies, and craniosynostosis.<sup>4-10</sup> But most commonly, the utility of these techniques has been studied in adult vertebral degenerative changes, mainly comparing the diagnostic performance of UTE and GRE bone imaging.<sup>11-14</sup> These studies in adult vertebral disorders have shown acceptable results with good agreement with CT or histology as the standard in the morphologic and quantitative assessment of the adult spine. There is not much literature on MR bone imaging in pediatric spinal disorders. We describe our experience with a novel MR bone imaging technique called fast field echo resembling a CT using restricted echo spacing (FRACTURE) MR in complex pediatric spinal anomalies. FRACTURE MR is a high-resolution 3D GRE technique that uses multiple constantly spaced echoes at the in phase of a 1.5T or 3T magnet.<sup>10,15</sup> Our study aimed to determine the reliability of FRACTURE MR for bone details in pediatric spinal anomalies as compared with the standard CT and to see if it can provide the intuitive CT-like bone appearance that is preferred by radiologists and spine surgeons.

## MATERIALS AND METHODS

### Study Design

The study was carried out prospectively after receiving approval from the institutional ethics board. The patient or their guardians were informed about the additional MR sequence that would be performed in imaging the child in addition to conventional MR and CT. An informed consent was obtained from them as per the consent forms approved by the ethics committee. The STROBE checklist was used to ensure completeness of this study.

### Patient Characteristics

Inclusion criteria included: 1) younger than 18 years of age, 2) referral for both MRI and CT imaging of the spine, and 3) no contraindication to MRI or sedation.

Exclusion criteria included: 1) patients older than 18 years of age, 2) only 1 of the 2 imaging modalities from MR and CT requested by the referring team, 3) patients uncooperative for imaging without sedation but have a contraindication to sedation, and 4) postoperative patients with metallic implants.

### Imaging Technique

MRI was performed on a 3T Ingenia (Philips Healthcare) by using a pediatric spine coil. FRACTURE sequence was acquired as a part of the routine conventional MR spine protocol and was accelerated by using compressed sensitivity encoding (CS). Parameters of the FRACTURE MR sequence are listed in Table 1. A CS factor of 8 was used to reduce the scan time. CT imaging was performed on a 64-slice Siemens scanner with the following protocol parameters: 80 kV, 130 mAs, 1-mm slice thickness utilizing bone-specific kernel.

### Postprocessing

This step includes a summation of the magnitude of all echoes to increase the SNR and then subtracting the image formed by the last echo from the summation to invert the gray scale and give the bone a CT-like image. The resultant image has better contrast between bone, bone marrow, and surrounding tissues.<sup>15</sup>

**Table 1: Parameter settings of the FRACTURE MR sequence**

Number of slices	80
Slice gap (mm)	0.6–1 mm
Sensitivity encoding	Yes
Stacks	1
Fold over suppression	Oversampling
Flip angle (°)	8
TR (ms)	11.2
TE (ms) (in-phase)	1.8
Echo spacing (ms)	2.3
Shim	Default

### **Image Analysis**

Both the CT and FRACTURE MR images were analyzed on the Osirix DICOM image viewer (<http://www.osirix-viewer.com>) independently by 2 radiologists, 1 with 13 years and another with 3 years of experience in radiology. All the studies of both modalities were anonymized by using the Osirix tool. Different parameters like fusion of the vertebral bodies, posterior elements, presence of anomalous vertebrae (like hemi-vertebra, flattened, block, or butterfly vertebra, endplate irregularities), and alignment deformity of the spine were studied and described. These parameters were selected based on the rationale that these bony abnormalities need to be assessed or ruled out on imaging to determine the extent of spinal deformity and formulate a definitive surgical plan tailored for each patient. These parameters were first evaluated on FRACTURE MR to remain blinded to the standard CT and after a gap of approximately 3 weeks to avoid memory bias, the same were assessed on CT. The quality of the FRACTURE images was graded on a 4-point Likert scale considering CT as the standard as 1) nondiagnostic, 2) artifacts causing diagnostic dilemma, 3) minimal artifacts not affecting diagnosis, and 4) excellent image quality with no artifacts. In addition, the SNR was calculated by ROIs placed on the vertebral body and in the background air for noise on FRACTURE MR and then by using the ratio of mean signal intensity of the bone to standard deviation of the background noise. Similarly, the contrast between the cortical and cancellous bone was calculated by using the ratio of the difference between the mean signal intensity of the cortical and cancellous bone to the SD of the background noise.<sup>14</sup>

### **Data Analysis**

For quantitative analysis of the descriptive data, the findings of both the radiologists on FRACTURE MR and CT were coded and compared to calculate interrater and intermodality agreement. Minor description variabilities about the deformity were disregarded but major differences like missed or overcalled deformities were considered as disagreement.

### **Statistical Analysis**

All the data were collected by using Excel (Microsoft) and statistical analysis was performed by the SPSS program for Windows, Version 28.0 (IBM). Agreement between 2 radiologists' opinions was assessed by using Cohen  $\kappa$  value and compared by using Fisher exact test.

## **RESULTS**

Eleven children with congenital spinal deformities ranging from Klippel Fiel (KF) anomaly ( $n = 6$ ), thoracolumbar vertebral formation-segmentation anomalies ( $n = 1$ ), and spinal dysraphism ( $n = 4$ ) were included in this study. The youngest patient was 1 month old and the oldest was 13 years old with 8 girls and 3 boys in the study sample.

### **Image Quality**

The average Likert score of the ratings of all patients by both radiologists was 3.4 for the FRACTURE MR image quality. Minimal artifacts were seen in 3 patients and artifacts limiting the diagnostic quality of FRACTURE MR images were seen in 1 patient; in

the remainder of the 7 patients the image quality was optimal. The total scan time for FRACTURE MR varied from 6–8 minutes that was reduced to 2–3 minutes by using CS.

Considering the differences in the FOVs, matrix size, and slice thickness in different patients, the lowest SNR was 70 and the highest was 150. The contrast of the cortical to cancellous bone was calculated for 5 children who were older than 5 years of age and the average was 19.2, the lowest being 19 and the highest value being 21. The remainder of the 6 children were younger than 3 years of age and showed immature vertebrae and lacked a well-formed cortex. The CT images in these younger patients also lacked a mature cortex.

### **Tissue Signal**

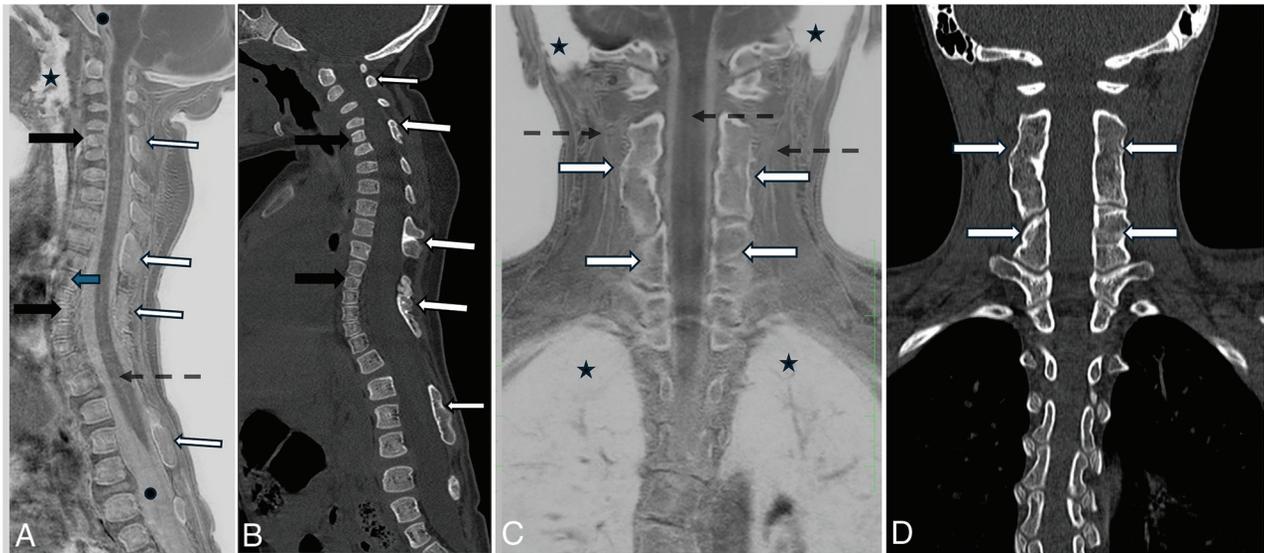
The different tissues in the anatomic region of interest on FRACTURE MR images appear as black, white, or in shades of gray. For example, the air in the airway and the lungs appear bright white, and CSF appears bright but less than that of the air signal. Soft tissues like the spinal cord, muscles, solid viscera, cerebellum, and the brainstem appear gray. Rapidly flowing blood in the arteries and the pulmonary vessels appears as a dark signal void similar to their signal on conventional MR. The vertebrae, ileum, sternum, ribs, and all other bones show differential signals for the cortical and trabecular bone. These normal tissue signal intensities have been demonstrated in the images in Figs 1–4. In addition, in 2 of our patients with lipomyelocoele and subcutaneous lipoma, there was an anomalous soft tissue that mimicked bone on FRACTURE MR images but lacked ossification on CT, which we presumed represented cartilaginous dysraphic hamartoma with microscopic mineralization, but the pathology diagnosis could not be traced because the patients were lost to follow-up (Supplemental Data).

### **Deformity Variables**

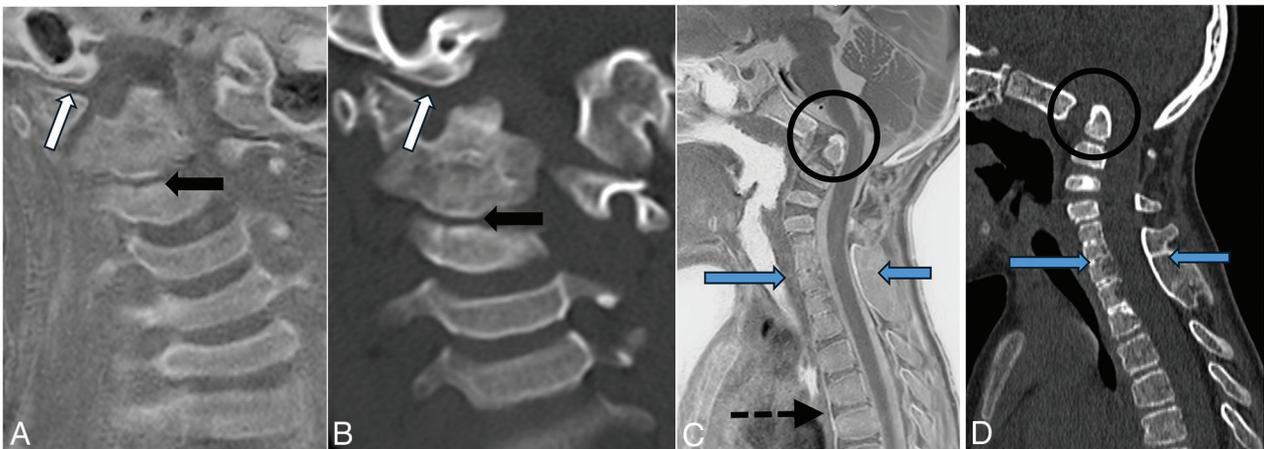
**Vertebral Body Fusion.** FRACTURE MR performed excellently in assessing the fusion of vertebral bodies with perfect interrater agreement. In 1 patient with multiple cervical vertebral fusion, the extent of fusion was overcalled and the extent of abnormality also varied in the FRACTURE MR findings of both the readers (Supplemental Data).

**Posterior Elements.** The results in the assessment of posterior element segmentation or formation anomalies were not consistent. The fusion was underestimated and varied between both readers on FRACTURE MR in 1 patient with segmentation anomaly and kyphoscoliotic thoracolumbar spine. The extent of fusion was overcalled by the readers in 1 patient each with right lateral lipomyelomeningocele (Supplemental Data) and KF anomaly. The second reader underestimated the fusion in a patient with a unilateral segmentation anomaly. Overall, there was substantial to nearing perfect interrater and intermodality agreement.

**Anomalous Vertebral Formation.** FRACTURE MR performed excellently in determining the anomalous formation of the vertebra. Both readers missed mild butterfly deformity with a very faint midline cleft in the cervical spine. In another patient, the artifacts across the vertebra at the level of upper and mid thoracic spine were called abnormal vertebra. Reader 1 interpreted this artifact along the superior endplate of T7 as subtle endplate



**FIG 1.** A, Sagittal CS-FRACTURE MR and (B) CT bone window images of the whole spine of a 2-year-old with KF anomaly show multilevel partial fusion of the cervical and thoracic vertebral bodies (*black arrows*) and fusion of posterior elements (*white arrows*). The bone details are well seen on the FRACTURE MRI, but lack of cortical maturation can be seen on images of both the modalities. There are artifacts along the fused lower thoracic vertebrae from adjacent cardiac motion (*blue arrow* in A). C, Coronal CS-FRACTURE MR and (D) CT bone window image of a 13-year-old with KF anomaly show fusion of multiple cervical vertebrae and their posterior elements (*white arrows*). Also note various normal signal intensities, very bright signal of air in the upper airway, lungs, and mastoid air cells (*black star* in A and C, respectively). CSF in the thecal space appears less bright than air (*black dot* in A); cord and other paraspinal soft tissues appear gray (*dotted black arrows* in A and C).



**FIG 2.** A, Coronal CS FRACTURE MR mean intensity projection and (B) CT bone window images of a 5-year-old with KF anomaly show right occipital condyle hypoplasia (*white arrows*), partial fusion of C2–C3 vertebrae (*black arrows*) and their posterior elements (not shown) resulting in left torticollis. C, Sagittal CS-FRACTURE MR and (D) CT bone window in a 6-year-old with KF anomaly show fusion of multiple cervical vertebrae (*blue arrows*), atlanto-occipital assimilation, hypoplastic dens, and basilar invagination (*black circle*). In addition, the acute angulation and effacement of the anterior CSF space at the cervicomedullary junction caused by the changes at the craniovertebral junction are well seen on FRACTURE MRI. The ligaments show a similar bright signal as CSF (*dotted black arrow* in C for anterior longitudinal ligament).

erosion on a background of endplate erosions at a few other levels in the same patient, whereas reader 2 interpreted a horizontally oriented artifact as a cleft in the T7 vertebra. Overall, the inter-rater and intermodality agreement was substantial to perfect.

**Alignment Deformity.** FRACTURE MR images provided optimal details of the alignment deformities like kyphosis, scoliosis, torticollis, and basilar invagination due to their ability to reformat with perfect interrater and intertechnique agreement. Disagreement in 1 patient was due to interrater variability in the description of the cause of left torticollis.

The results of the interreader and intertechnique agreement for different parameters are described in Table 2.

## DISCUSSION

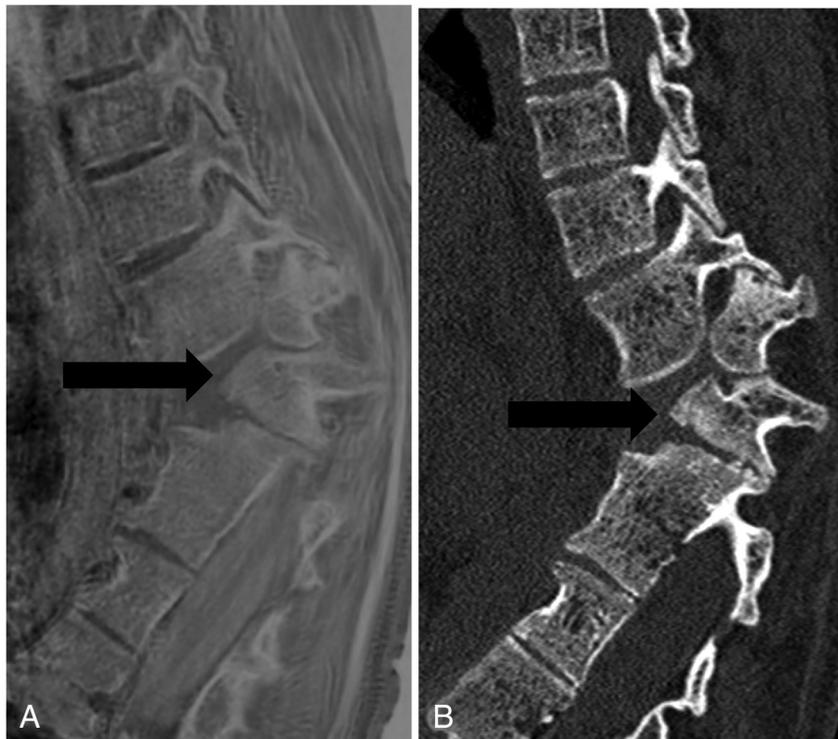
In this study, we evaluated the feasibility of the FRACTURE MR images in pediatric congenital spinal anomalies. The quality of the FRACTURE MR images was acceptable in most of our patients with diagnostically limiting artifacts in 1 patient and minimal artifacts in 3 patients. The artifacts in these patients arose from different factors, for example, in a patient with a false-positive

vertebral anomaly on a coronal image acquisition of FRACTURE MR, the horizontally oriented artifact was believed to be from the motion of pulsating aortic arch branches or breathing.

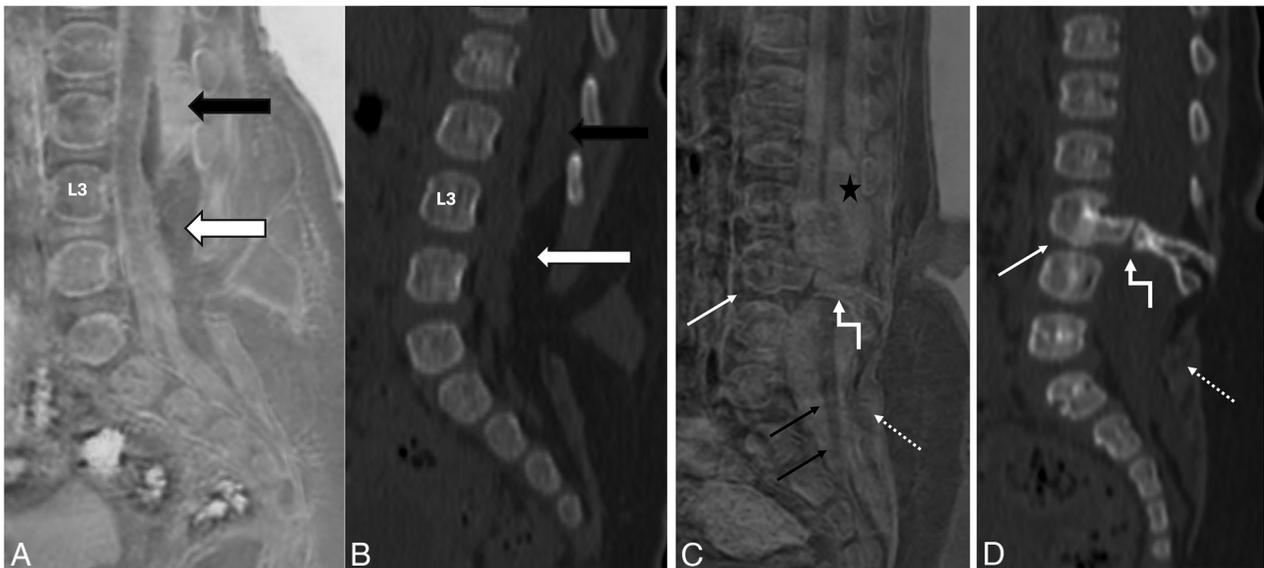
Interestingly, this artifact was also seen on CT images at the same level and was believed to be due to breathing artifacts. In another patient with thoracic lordosis, the artifacts were minimal but seemed to be a result of cardiac motion and its proximity to the spine due to thoracic lordosis. In the other 2 patients, the artifacts were from mild motion resulting from the fading effect of sedation. It has been described that GRE, UTE, and ZTE techniques are all prone to motion and susceptibility from air in the lungs or metallic implants in the vicinity. However, UTE is known to be more prone to pulsation and motion, and GRE is more likely to be affected by susceptibility artifacts. ZTE is believed to be robust to motion due to its rapid *k*-space sampling.<sup>3,13,14</sup>

The SNR of the bone was in the acceptable range but lower in children younger than 3 years of age due to their immature cartilaginous skeleton. There was a good contrast of the cortical to the trabecular bone in all patients older than 5 years of age with relatively mature skeletons than in the younger children with more cartilaginous skeletons. (Fig 1A, 4A and C, Supplemental Data). However, the lower bone signal in younger children was not diagnostically

acceptable range but lower in children younger than 3 years of age due to their immature cartilaginous skeleton. There was a good contrast of the cortical to the trabecular bone in all patients older than 5 years of age with relatively mature skeletons than in the younger children with more cartilaginous skeletons. (Fig 1A, 4A and C, Supplemental Data). However, the lower bone signal in younger children was not diagnostically



**FIG 3.** A, Sagittal CS-FRACTURE MR and (B) CT of the thoracolumbar junction of a 12-year-old show complex fusion segmentation anomaly in the form of block and hemivertebrae resulting in a sharp kyphotic deformity (*black arrows*).



**FIG 4.** A, Sagittal CS-FRACTURE MR and (B) CT bone window images in a 6-month-old show defect in the posterior elements from L3 level down with a closed spina bifida deformity. In addition, there is an arachnoid cyst (*black arrows*) that shows CSF-like signal on FRACTURE MRI. The cord is seen low lying and lipoma-neural placode interface lies within the spinal canal suggesting a lipomyelocele (*white arrows*). C, Sagittal CS-FRACTURE and (D) CT bone window images in a 6-month-old show partial fusion of L3 and L4 vertebral bodies (*white solid arrows*), posterior element defects from L4 below (*dotted white arrows*). An osseous bar (*elbow arrows*) is seen at L3–L4 level resulting in hemi cords suggestive of Pang type 1 diastematomyelia. Syrinx (*star*) and low-lying cord (*black arrows*) are well seen on FRACTURE images. Also note the low bone signal and lack of cortical maturation in both the patients due their very young age and cartilaginous vertebrae.

**Table 2: Analysis of the intertechnique and interrater agreement**

Parameter	Radiologist 1			Radiologist 2			Interreader Reliability (%)
	CT-MR Agreement (%)	Sensitivity (%)	Specificity (%)	CT-MR Agreement (%)	Sensitivity (%)	Specificity (%)	
	Vertebral body fusion	91	100	75	91	100	
Posterior elements fusion	82	89	50	64	78	50	64
Anomalous vertebral shape	82	80	82	82	80	83	91
Alignment	100	100	100	100	100	100	91

limiting. The agreement between the 2 modalities for each reader was perfect in assessing vertebral fusion, anomalous shape, and alignment deformity but was moderate to substantial for posterior element segmentation and formation anomaly. The misinterpretations of posterior elements were mainly in children with complex spinal anomalies with multilevel fusion and formation anomalies. We believe that the small size and abnormal orientation of the dysplastic posterior elements on a background of paraspinal soft tissue details results in a diagnostic dilemma. Most studies on GRE bone imaging in adult vertebral pathologies focus on evaluating vertebral endplate changes, neural foramen stenosis, or spondylolysis. However, these studies do not address pathologies necessitating detailed assessment of the posterior elements, which is primarily required in pediatric congenital vertebral anomalies. These studies describe GRE images to have a higher spatial resolution, better bone-to-soft tissue contrast, and sharper images as compared with UTE.<sup>3,11,13</sup> UTE/ZTE MR bone images have a flatter contrast of the soft tissues making bone stand out as bright structures on a background of uniform gray signal of the surrounding soft tissue structures.<sup>14</sup>

Special considerations in pediatric MR bone imaging include the need for sedation, smaller anatomy, and immature ossification of the skeleton in neonates and early childhood, which can potentially pose challenges in obtaining optimal signal from osseous structures. Our experience with FRACTURE MR showed that it is feasible to use this novel MR technique in pediatric spine especially in older children (older than 3 years age) with better ossified vertebrae as compared with the very young. The sedation needed for the added sequence can be covered under the same sedation of conventional MR sequences or may require tailoring for the length of the study; this varies among children. The images obtained have optimal bone contrast with good resolution. But there are some challenges with GRE bone images like FRACTURE MR that have been described with adult vertebral imaging as well. Ligaments, tendons, fascia, air, and edema all appear white on these GRE-based MR bone images, which can cause diagnostic errors and interfere with bone segmentation (demonstrated in Fig 1A and 2C). Based on this, we can presume that marrow edema due to subtle vertebral fractures, trabecular injuries, or infection can be missed as it will appear white on these inverted scale images and can alter management. This becomes particularly important in cases of suspected nonaccidental injuries. In the use of these sequences in adult vertebral pathologies, it has been suggested that conventional fluid-sensitive MR images can be helpful in such cases to identify areas of bone marrow edema.<sup>3,13,16</sup> For diagnosis of true sclerosis, GRE images were found to be superior to UTE and SWI.<sup>17,18</sup> Calcification of

the ligaments and intervertebral air are not commonly encountered in pediatric spine pathologies but these factors in general become important if we intend to use these in pediatric musculoskeletal imaging and head injuries. In head trauma, we presume that the fractures in the vicinity of the pneumatized bones like paranasal sinuses and mastoids can be missed on FRACTURE MR as has been described with black bone imaging due to similar signals of air and bone.<sup>5</sup>

Of the different techniques available for bone imaging, FRACTURE MR holds many advantages. Being a 3D GRE technique-based sequence, it is widely available on most of the commonly used scanners and is also independent of the magnetic field strength, allowing wide-scale applicability. Second, the images obtained have high spatial resolution with good bone-to-soft tissue contrast. Also, the images obtained are 3D, making reformations possible. Third, the postprocessing is simple and minimal to obtain CT-like images.<sup>15</sup> On the other hand, techniques like UTE and ZTE have high-end software/hardware requirements and provide CT-like images of lesser contrast-to-noise ratio compared with the GRE images. They also have spatially nonselective excitations that induce large FOVs and reduce their flexibility with respect to other parameters. Hence not much trade-off between the spatial resolution, FOV, and time for imaging can be achieved with this sequence. Susceptibility-weighted bone imaging, though easily accessible with minimal postprocessing, is dependent on field strength.<sup>3,11,13,19</sup> A negative correlation was found between the sensitivity of SWI black bone imaging and magnetic field strength in pediatric head imaging for traumatic brain injuries.<sup>5</sup>

It is worthwhile to mention that there are some limitations to the use of FRACTURE MR in routine practice across different age groups and indications. First, is the scan time, which can be up to 6–8 minutes in pediatric spine and even higher in larger FOVs like adult spines. This drawback is more or less common to all MR bone imaging sequences. The scan time can be reduced by nearly one-half by using parallel imaging techniques like CS as in our case. This duration is still longer than the acquisition time of the spine CT, which is a matter of a few seconds. This increases the possibility of motion artifacts in nonsedated children and makes it an impractical choice in patients with acute trauma. The other drawback is its limited use in postoperative patients with metallic implants due to susceptibility artifacts. Third, compatibility of FRACTURE MR images with neuronavigation systems is not yet established. Synthetic CTs and artificial intelligence-guided postprocessing techniques might help in the future in achieving these shortcomings.<sup>20,21</sup>

There are also some limitations in our study that we acknowledge. First, the cohort of pediatric patients is small and from a single center, thereby not considering the technical differences

between the studies. Second, we have focused only on congenital spinal anomalies and have not studied their use in pediatric spine trauma, vertebral tumors, or osteomyelitis. Third, we did not compare the performance of FRACTURE MR with the conventional MR sequences or other MR bone imaging techniques.

## CONCLUSIONS

FRACTURE MR is reliable for bone imaging in pediatric complex spinal anomalies with substantial to perfect agreement between the 2 modalities. The bone signal is lower in children younger than 3 years due to a more cartilaginous skeleton. Despite some drawbacks of the sequences, it has the potential to complement conventional MR sequences in place of CT for bone assessment in pediatric complex spinal anomalies. This can reduce the radiation dose and an additional study in this group of pediatric patients who will be serially followed up with imaging during their management. We believe that FRACTURE MR still in its naïve form cannot replace CT, at least in trauma due to time constraints and postoperative studies due to susceptibility from metallic implants. Whether it holds potential in pediatric spine trauma or other vertebral pathologies needs to be evaluated by additional studies.

## ACKNOWLEDGMENTS

We acknowledge the contribution of Walmik Mistri in statistical analysis of the data.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Trenga AP, Singla A, Feger MA, et al. **Patterns of congenital bony spinal deformity and associated neural anomalies on X-ray and magnetic resonance imaging.** *J Child Orthop* 2016;10:343–52 CrossRef Medline
2. Prahinski JR, Polly DW, McHale KA, et al. **Occult intraspinal anomalies in congenital scoliosis.** *J Pediatr Orthop* 2000;20:59–63 Medline
3. Florkow MC, Willemsen K, Mascarenhas VV, et al. **Magnetic resonance imaging versus computed tomography for three-dimensional bone imaging of musculoskeletal pathologies: a review.** *J Magn Reson Imaging* 2022;56:11–34 CrossRef Medline
4. Cho SB, Baek HJ, Ryu KH, et al. **Clinical feasibility of zero TE skull MRI in patients with head trauma in comparison with CT: a single-center study.** *AJNR Am J Neuroradiol* 2019;40:109–15 CrossRef Medline
5. Dremmen MHG, Wagner MW, Bosemani T, et al. **Does the addition of a “black bone” sequence to a fast multisequence trauma MR protocol allow MRI to replace CT after traumatic brain injury in children?** *AJNR Am J Neuroradiol* 2017;38:2187–92 CrossRef Medline
6. Kralik SF, Supakul N, Wu IC, et al. **Black bone MRI with 3D reconstruction for the detection of skull fractures in children with**

- suspected abusive head trauma.** *Neuroradiology* 2019;61:81–87 CrossRef Medline
7. Sandberg JK, Young VA, Yuan J, et al. **Zero echo time pediatric musculoskeletal magnetic resonance imaging: initial experience.** *Pediatr Radiol* 2021;51:2549–60 CrossRef Medline
8. Gersing AS, Pfeiffer D, Kopp FK, et al. **Evaluation of MR-derived CT-like images and simulated radiographs compared to conventional radiography in patients with benign and malignant bone tumors.** *Eur Radiol* 2019;29:13–21 CrossRef Medline
9. Low XZ, Lim MC, Nga V, et al. **Clinical application of “black bone” imaging in paediatric craniofacial disorders.** *Br J Radiol* 2021;94:20200061 CrossRef Medline
10. Leonhardt Y, Kronthaler S, Feuerriegel G, et al. **CT-like MR-derived images for the assessment of craniosynostosis and other pathologies of the pediatric skull.** *Clin Neuroradiol* 2023;33:57–64 CrossRef Medline
11. Deininger-Czermak E, Gascho D, Franckenberg S, et al. **Added value of ultra-short echo time and fast field echo using restricted echo-spacing MR imaging in the assessment of the osseous cervical spine.** *Radiol Med* 2023;128:234–41 CrossRef Medline
12. Gassert FT, Kufner A, Renz M, et al. **Comparing CT-like images based on ultra-short echo time and gradient echo T1-weighted MRI sequences for the assessment of vertebral disorders using histology and true CT as the reference standard.** *J Magn Reson Imaging* 2024;59:1542–52 CrossRef Medline
13. Schwaiger BJ, Schneider C, Kronthaler S, et al. **CT-like images based on T1 spoiled gradient-echo and ultra-short echo time MRI sequences for the assessment of vertebral fractures and degenerative bone changes of the spine.** *Eur Radiol* 2021;31:4680–89 CrossRef Medline
14. Vucevic D, Malis V, Yamashita Y, et al. **Ultrashort echo time and fast field echo imaging for spine bone imaging with application in spondylolysis evaluation.** *Computation* 2024;12:152 CrossRef
15. Johnson B, Alizai H, Dempsey M. **Fast field echo resembling a CT using restricted echo-spacing (FRACTURE): a novel MRI technique with superior bone contrast.** *Skeletal Radiol* 2021;50:1705–13 CrossRef Medline
16. Gascho D, Zoelch N, Tappero C, et al. **FRACTURE MRI: optimized 3D multi-echo in-phase sequence for bone damage assessment in craniocerebral gunshot injuries.** *Diagn Interv Imaging* 2020;101:611–15 CrossRef Medline
17. Diekhoff T, Hermann KGA, Greese J, et al. **Comparison of MRI with radiography for detecting structural lesions of the sacroiliac joint using CT as standard of reference: results from the SIMACT study.** *Ann Rheum Dis* 2017;76:1502–08 CrossRef Medline
18. Diekhoff T, Greese J, Sieper J, et al. **Improved detection of erosions in the sacroiliac joints on MRI with volumetric interpolated breath-hold examination (VIBE): results from the SIMACT study.** *Ann Rheum Dis* 2018;77:1585–89 CrossRef Medline
19. Tsuchiya K, Gomyo M, Katase S, et al. **Magnetic resonance bone imaging: applications to vertebral lesions.** *Jpn J Radiol* 2023;41:1173–85 CrossRef Medline
20. Lombardi AF, Ma YJ, Jang H, et al. **Synthetic CT in musculoskeletal disorders: a systematic review.** *Invest Radiol* 2023;58:43–59 CrossRef Medline
21. Chong LR, Lee K, Sim FY. **3D MRI with CT-like bone contrast—An overview of current approaches and practical clinical implementation.** *Eur J Radiol* 2021;143:109915 CrossRef Medline

# Management Outcomes after Image-Guided Percutaneous Biopsy for Suspected Vertebral Osteomyelitis-Discitis

Dania G. Malik,  Mathew V. Smith, Gabriel M. Swenson, Robert E. Grady,  Felix E. Diehn,  Erik H. Middlebrooks,  Jeffrey S. Ross, and  Tanya J. Rath



## ABSTRACT

**BACKGROUND AND PURPOSE:** Studies show a modest yield for image-guided biopsy of suspected vertebral osteomyelitis-discitis. Many studies evaluate factors to improve diagnostic yield, and few studies assess how biopsy results impact clinical management. We aim to evaluate the impact of biopsy results on clinical management in suspected vertebral osteomyelitis-discitis.

**MATERIALS AND METHODS:** We performed a retrospective study of patients who underwent image-guided biopsy for suspected vertebral osteomyelitis-discitis. Data collected included risk factors, imaging findings, laboratory values, antibiotics, biopsy procedure details, microbiology and pathology results, and clinical course. Factors assessed for management change included whether biopsy results affected antibiotic type or course, decision to start or stop antibiotics, surgical decisions, or if an alternate diagnosis was determined.

**RESULTS:** Three hundred ten biopsies were included. Biopsy yield with true-positive culture results was 34% (104/310) and similar for patients on antibiotics (36%, 34/94) and off antibiotics (32%, 66/204). Yield was greater when disc was sampled (36%, 82/228) versus bone only (8%, 2/24) and with aspiration of disc and/or bone (42%, 39/92) versus core only (29%, 56/193). With positive blood cultures before biopsy, biopsy yield was 50% (22/44) with concordance and discordance rates of 75% (18/24) and 17% (4/24), respectively, and 8% (2/24) of positive biopsy results deemed contaminants. Management was affected in 36% (113/310) of all biopsies and in 78% (81/104) of biopsies with a positive culture result. No management change occurred in 57% (177/310) of biopsies. Management change was unclear in 6% (20/310). Biopsy results changed antibiotics in 27% (85/310). Management change occurred in 23% (10/44) of cases with prior positive blood culture compared with 41% (93/233) without a prior culture source ( $P = .024$ ). Negative culture results influenced management in 16% (32/194).

**CONCLUSIONS:** Image-guided biopsy for vertebral osteomyelitis-discitis has a meaningful impact on management despite modest yield. Greatest management impact is seen with positive culture results, no prior culture source, and patients not on antibiotics at the time of biopsy. Biopsy culture yield is not affected by preceding antibiotics, and yield is greater with disc sampling and aspiration.

**ABBREVIATIONS:** CRP = C-reactive protein; ESR = erythrocyte sedimentation rate; IDSA = Infectious Disease Society of America; MC = management change; NMC = no management change; UC = unclear; VO = vertebral osteomyelitis-discitis; WBC = white blood cell

Vertebral osteomyelitis-discitis (VO) is an infection of the intervertebral disc and vertebrae, usually acquired hematogenously. *S aureus* is the predominant pathogen in Western countries, and *M tuberculosis* is the most common worldwide.<sup>1</sup> Symptoms are vague, most commonly include pain, and progress insidiously over weeks.<sup>1,2</sup> Laboratory assays and imaging supplement clinical suspicion to establish the diagnosis. When no

organism is identified by blood culture, image-guided biopsy is performed to identify a causative organism.<sup>2</sup> In patients with high clinical suspicion, empiric antimicrobial therapy is often initiated before microbiologic diagnosis.<sup>2</sup> Regardless of antimicrobial treatment, the yield of a percutaneous biopsy is modest,<sup>3-5</sup> with positive microbiology results in 27%<sup>5</sup> and biopsy sensitivities of 46%–52%.<sup>5,6</sup>

The limited yield of image-guided biopsy and the initiation of empiric antibiotics before microbiologic diagnosis raise consideration of the value of biopsy and how to optimize it. A study on image-guided biopsies for suspected osteomyelitis throughout the body found biopsy cultures influencing treatment in 9%.<sup>7</sup> In studies evaluating biopsies for VO, reported management impact ranges from 9%–50%.<sup>3,8-10</sup>

Received September 25, 2024; accepted after revision December 18.

From the Department of Radiology (D.G.M., M.V.S., G.M.S., R.E.G., J.S.R., T.J.R.), Mayo Clinic, Phoenix, Arizona; Department of Radiology (F.E.D.), Mayo Clinic, Rochester, Minnesota; and Department of Radiology (E.H.M.), Mayo Clinic, Jacksonville, Florida.

Please address correspondence to Dania G. Malik, 5777 E Mayo Blvd, Phoenix, AZ 85054; e-mail: Malik.dania@mayo.edu

 Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8630>

## SUMMARY

**PREVIOUS LITERATURE:** Image-guided biopsy for diagnosis of VO has been studied to evaluate factors to improve diagnostic yield. Image-guided biopsy has modest yield (27%) and sensitivity (46%-52%). Studies report biopsy influences management of VO in 9%–50% of cases with varying results, especially in the setting of culture-negative cases.

**KEY FINDINGS:** Biopsy results changed management in 36% of cases, with 78% of positive cultures impacting antibiotic therapy. Disc sampling and aspiration yielded higher positive culture results. Antibiotics before biopsy did not impact yield, but management change was greater in patients not on antibiotics and with no prior culture source.

**KNOWLEDGE ADVANCEMENT:** This study shows that image-guided biopsy meaningfully impacts management of VO, especially when no prior organism is identified and patients are not on antibiotics. We highlight that the biopsy yield is improved with disc sampling and aspiration and that the yield is not significantly affected by prior antibiotic use.

We aim to expand upon prior studies to evaluate the yield and therapeutic impact of image-guided biopsy performed for suspected VO and identify which patients may or may not benefit from biopsy.

## MATERIALS AND METHODS

Following Institutional Review Board approval, we performed a multisite retrospective review of adult patients who underwent image-guided biopsy for clinically suspected VO from February 1998 to July 2023.

We searched radiology and pathology reports in the Illuminate InSight software with keywords including spinal levels (“C2,” “T1,” etc), “biopsy,” and “osteomyelitis.” CT and fluoroscopic biopsies for suspected VO were included, while biopsies for non-infectious processes were excluded. Patient demographics, medical history, imaging findings, laboratory values, biopsy details, microbiology and pathology results, antibiotic regimen, clinical course, and management changes were recorded from the medical record. Management change factors assessed included biopsy results’ impact on antibiotic type or course, initiating or stopping antibiotics, surgical decisions, or determining alternate diagnoses.

Over the 25-year period, different tissue processing methods, media, and vendors were used across the sites. All employed routine era-appropriate culture techniques. Biopsies were processed by using a stomacher machine or plastic tube grinder to homogenize along with sterile saline or broth if needed. Generally, aerobic cultures were incubated at 35°C in 5% CO<sub>2</sub> for 5 days, aerobic plates (if performed) incubated for at least 4 days, and thioglycollate broth incubated for 14 days. Starting in 2016, some homogenates were inoculated into blood culture bottles (BD BACTEC Plus Aerobic/F medium and BD BACTEC Lytic/10 Anaerobic/F medium) and incubated on the BACTEC 9240/FX instruments (BD Diagnostic Systems) for 14 days. Polymerase chain reaction was not routinely used, though it was available on request.

The  $\chi^2$  test was used to compare proportions between independent proportions, and the 2-sample t-test was used to compare means. Agreement between prior culture sources and biopsy culture results was evaluated by using Cohen  $\kappa$ . Statistical analysis was performed by using Excel (Microsoft), MedCalc for Windows, Version 19.4 (MedCalc Software), and R Version 4.2.2 by using package *irr*. All statistical tests were 2-sided, and a *P* value of < .05

was considered statistically significant. In subgroup analysis of variables, cases with missing data were excluded. This study follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines (Supplemental Data).

## RESULTS

Three hundred ten image-guided biopsies of disc, vertebra, and/or paraspinal tissue for suspicion of VO were included among 297 unique patients. Twelve patients had multiple biopsies: 9 patients underwent repeat biopsies for the same presentation, 2 patients had 2 separate presentations at different time points, and 1 patient had a repeat biopsy for the same presentation and another separate presentation at a different time point. Among the unique patients were 111 women and 186 men with a mean age of 65.55 years (SD 12.71).

### Imaging Findings

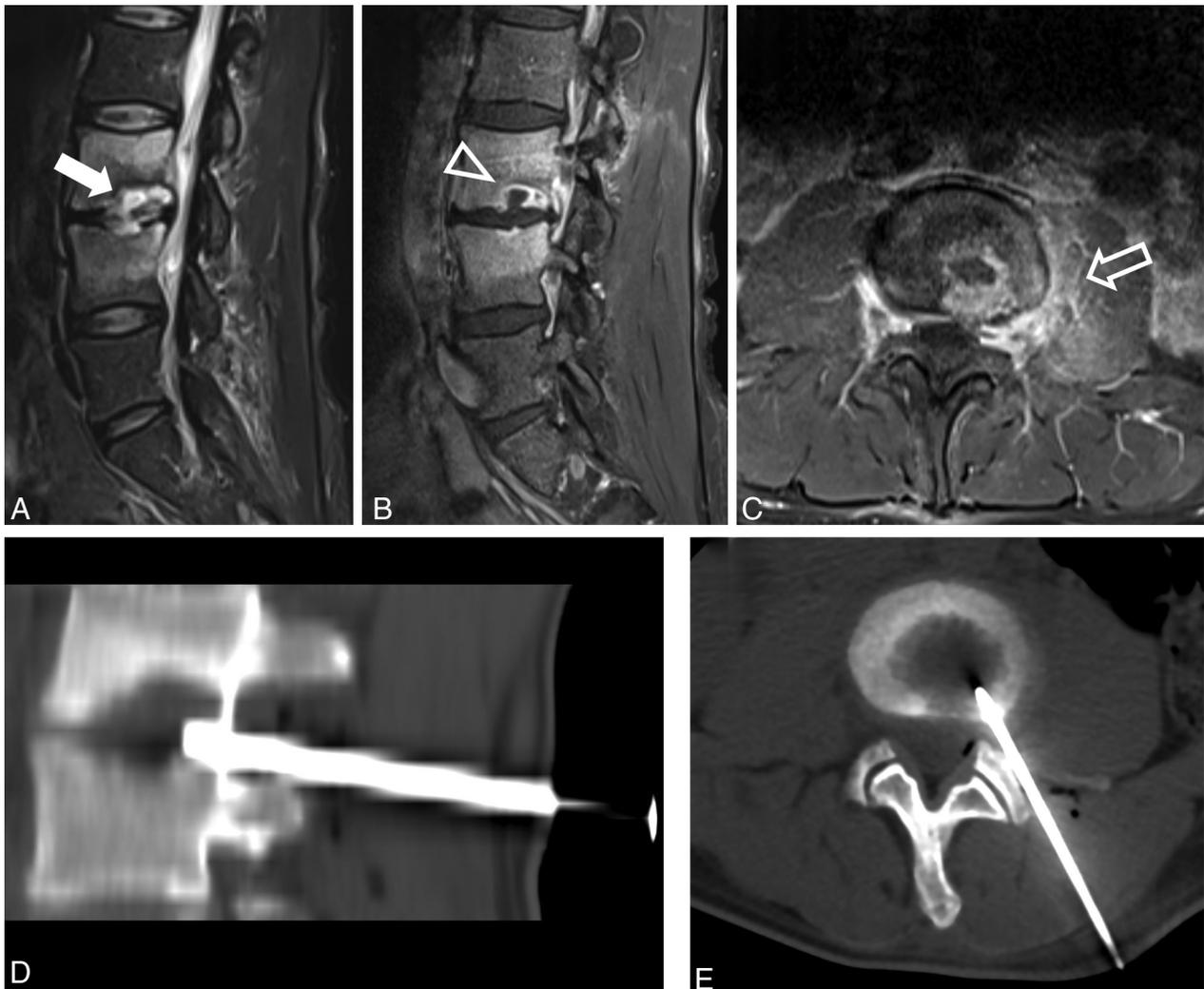
Imaging reports before biopsy were available for 94% (290/310): both MR and CT reports were available in 28% (88/290), MR only in 54% (167/290), and CT only in 11% (35/290). Seventy-nine percent (231/290) of reports favored VO, 14% (40/290) were equivocal, and 7% (19/290) favored noninfectious diagnoses. Biopsy was recommended in 5% of reports (14/290). Figure 1 shows imaging of a true-positive VO case.

### Lab Values

White blood cell (WBC) count was available for 255 patients, with an average of  $8.5 \times 10^9/L$  (SD 7.69) within normal range ( $4\text{--}11 \times 10^9/L$ ). WBC was abnormal in 22% (55/255), with 39 having elevated WBC and 16 having low WBC. Erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) were available in 226 and 239 cases, respectively. Average ESR was 46.6 mm/h (SD 31.5, normal <20 mm/h), with elevated ESR in 77% (174/226). Average CRP was 49.1 mg/dL (SD 60.2, normal <8 mg/L), with elevated CRP in 78% (187/239). Positive blood cultures were documented in 14% (44/310) before biopsy, and 11% (33/310) had other positive culture/infection sources (urine, joints, wound, sputum, serology, prior biopsy).

### Biopsy Procedure Details

Biopsy reports were available in 309 of 310 cases (details summarized in Table 1). CT and fluoroscopic guidance were used for



**FIG 1.** Sagittal STIR (A), sagittal postcontrast T1 (B), and axial postcontrast T1 (C) MR sequences demonstrate imaging of L3-L4 vertebral osteomyelitis-discitis with disc-endplate irregularity, T2 hyperintensity centered within the disc (white arrow) extending in the adjacent vertebral bodies, and peripheral enhancement of the irregular disc and endplates (outline arrowhead) with enhancement of the vertebral bodies and the paraspinal tissues (outline arrow). CT-guided biopsy with oblique sagittal (D) and axial (E) images show biopsy needle sampling of the disc-endplate margin.

53% (165/309) and 47% (144/309), respectively. Distribution by level was 7% (21/310) cervical, 33% (102/310) thoracic, and 60% (187/310) lumbar.

Needle sizes were reported in 245 cases ranging from 9–22 g. Many biopsies used multiple needle gauges, including the introducer, to collect samples. The most common (51%, 124/245) needle size was 14 g with 11 g/13 g introducers. Of the core biopsies, 68% (197/291) quantified sample number (range 1–13; mean 4.0 [SD 2.1]), and others reported “multiple” or “several” core samples without quantification. Sample length was not routinely reported. Aspiration volume was reported in 45% (52/116), ranging from 0.5–15 mL and a mean of 3.1 mL (SD 3.0).

Complications were documented in 1.6% (5/310), 3 with CT and 2 with fluoroscopic guidance, including prolonged bleeding ( $n = 1$ ), nausea ( $n = 1$ ), pain greater than expected ( $n = 2$ ), and a linear fracture through the adjacent facet joint ( $n = 1$ ).

#### Biopsy Yield

Biopsy yielded a positive culture result in 38% (119/310) of cases, with 13% (15/119) deemed contaminants by the clinical team,

resulting in a 34% (104/310) true-positive culture yield. Negative cultures were documented in 58% (179/310). Culture results were unavailable for 4% (12/310) because of lost specimens, nonsubmission, or patient death or discharge. Supplemental Data provides clinical and procedure information in relation to biopsy yield.

As shown in the Supplemental Data, there was no significant difference in biopsy yield with respect to technique, needle size, sample number, and aspirate volume. Biopsy needle sizes (9–13 g, 14 g, 15–17 g, 18–19 g, 20–22 g) showed no significant yield difference, with the lowest yield of 31% (38/124) in the 14 g group and highest yield of 38% (8/21) in the 15–17 g group ( $P = .6$ ). Core-only biopsies of bone and/or disc had significantly lower yield (29%, 58/193) compared with disc/bone aspirate with or without cores (42%, 39/92) ( $P = .002$ ). Aspiration only resulted in a yield of 56% (10/18). Bone only samples had significantly lower yield versus disc with or without bone (yield 36%, 82/228) ( $P = .006$ ). The yield for paraspinal tissue/fluid sampling was 43% (13/30) compared with 33% (84/252) without paraspinal sampling ( $P = .3$ ).

**Table 1: Biopsy procedure details**

Procedure Detail	Category	Value
Technique	CT	53% (165/309)
	Fluoroscopy	47% (144/309)
Level	Cervical	7% (21/310)
	Thoracic	33% (102/310)
	Lumbar	60% (87/310)
Tissue	Disc + bone	40% (123/309)
	Bone only	8% (24/309)
	Disc only	34% (105/309)
	Paraspinal + bone/disc	9% (29/309)
	Paraspinal only	0.3% (1/309)
	Unspecified	9% (27/309)
Sample	Core + aspirate	32% (98/309)
	Core only	62% (193/309)
	Aspirate only	6% (18/309)
Primary needle size	9–13 g	14% (42/309)
	14 g	40% (124/309)
	15–17 g	7% (21/309)
	18–19 g	13% (42/309)
	20–22 g	5% (16/309)
	Unspecified	21% (64/309)
	Core samples	Average number
Aspiration	Average volume	3.125 (SD 3.0)

Thirty percent (94/310) of patients were receiving antibiotics at the time of biopsy. Biopsy yield was 36% (34/94) on antibiotics versus 32% (66/204) not on antibiotics ( $P = .49$ ). Thirteen patients were on long-term antibiotics for prophylaxis or treatment of other infections, with 16% (2/13) having a positive culture result. Excluding long-term antibiotics ( $> 6$  weeks), mean antibiotic duration before biopsy was 3.69 days (SD 3.01) for positive cultures versus 6.35 days (SD 8.07) for negative cultures ( $P = .08$ ).

Seventy-seven cases had other positive culture sources, including 44 positive blood cultures. For positive blood cultures, 55% (24/44) had a positive biopsy culture, with 75% (18/24) growing the same organism, 17% (4/24) growing different organisms, and 8% (2/24) growing contaminant organisms. Cohen  $\kappa$  score for agreement between blood and biopsy cultures was 0.34 (95% CI:  $-0.02, 0.70$ ). With any prior positive culture result (blood, urine, wound, etc), 49% (38/77) had positive biopsy culture, with 68% (26/38) growing the same organism, 18% (7/38) growing different organisms, and 13% (5/38) growing contaminant organisms. The Cohen  $\kappa$  score for all prior cultures compared with biopsy results was 0.39 (95% CI: 0.02, 0.65). Positive biopsy yield was 50% (22/44) with prior positive blood cultures compared with 30% (71/233) with no prior culture source ( $P = .01$ ). Figure 2 lists the isolated organisms from the biopsy.

### Management Outcomes and Clinical Diagnosis

Biopsy results changed management (MC) in 36% (113/310) of all cases and did not change management (NMC) in 57% (177/310). Management impact was unclear (UC) in 6% (20/310) due to insufficient documentation or clinical complexity. Figure 3 categorizes management change related to biopsy results. Antibiotics were changed in 27% (85/310) of all cases. In positive biopsy cultures, management changed in 78% (81/104) when antibiotics were started or tailored based on the organism. Table 2 shows management change related to clinical variables.

For culture-negative or culture-contaminant cases, management changed in 16% (32/194), or 10% (32/310) of all cases, because of stopping antibiotics, not initiating antibiotics despite high clinical suspicion before biopsy, or determining another diagnosis. One case with negative culture had pathology consistent with acute-on-chronic inflammation, and antibiotics were started as a result of a presumed infection with false-negative culture. In these 32 cases, the sample was core only in 63% (20/32), aspirate only in 6% (2/32), and both core and aspirate in 31% (10/32). Sample location was bone only in 9% (3/32), disc only in 34% (11/32), bone and disc in 38% (12/32), paraspinal and bone/disc in 13% (4/32), and unspecified in 6% (2/32).

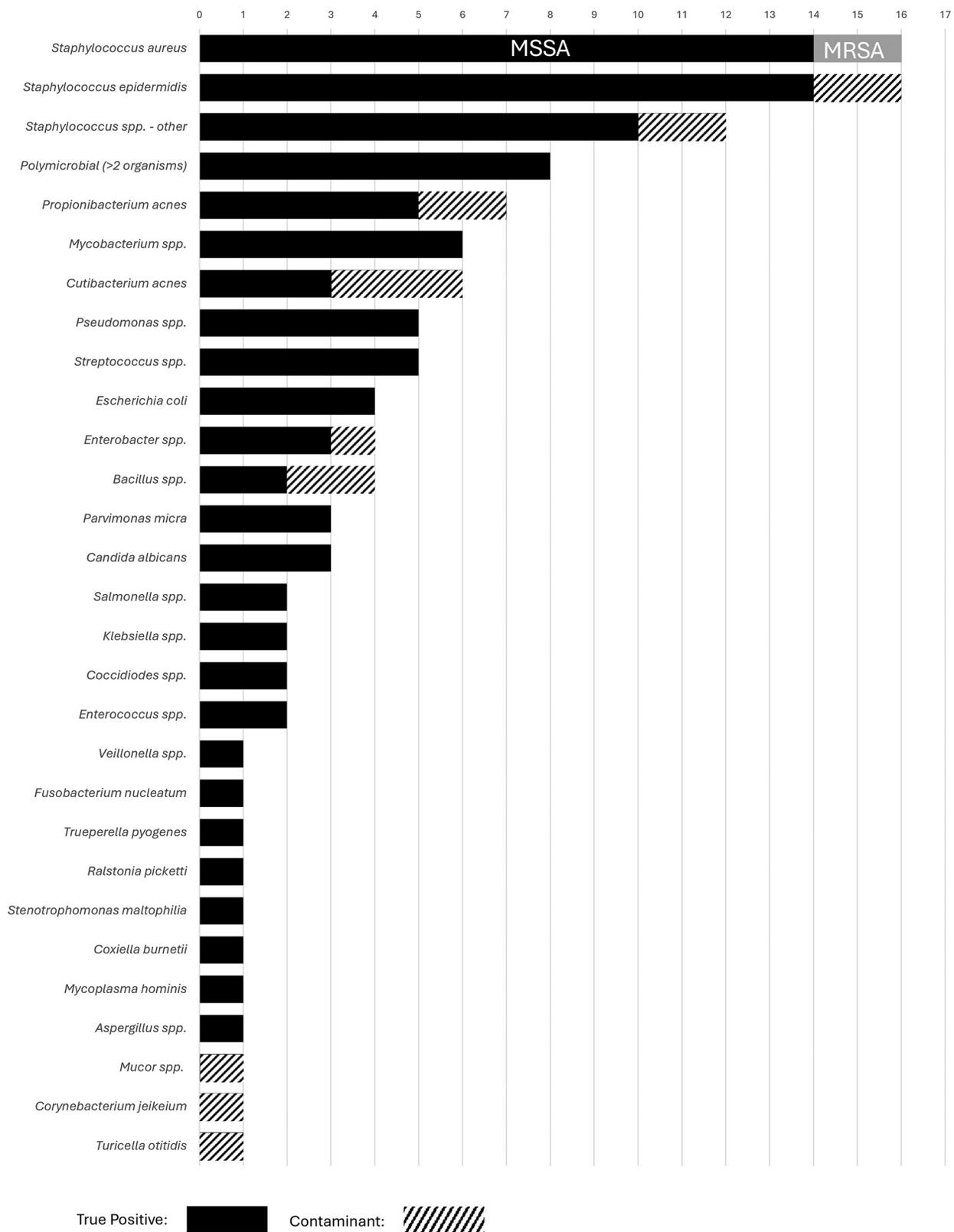
Biopsy results did not affect antibiotics in 57% (110/194) of culture-negative or culture-contaminant cases and 44% (134/310) of all cases treated with empiric antibiotics or antibiotics targeted to other culture sources. Antibiotics were not changed in 73% (32/44) of patients with positive blood cultures already on targeted therapy. Antibiotic changes with prior positive blood culture occurred with discordant biopsy results and inadequate antibiotic coverage (7%, 3/44), antibiotics held before biopsy with positive biopsy result (5%, 2/44), and broad-spectrum antibiotics narrowed after concordant or negative biopsy result (11%, 5/44). Excluding patients with no other prior culture source (Fig 4), management changed in 40% (93/233), significantly greater than 23% (10/44) with prior positive blood culture ( $P = .04$ ). In patients not on antibiotics at biopsy, management changed for 41% (88/216), significantly greater than 22% (21/94) on antibiotics ( $P = .001$ ).

Culture-negative and culture-contaminant cases were deemed to not contribute to management change if repeat biopsy or surgery was needed ( $n = 14$ ), in cases with low suspicion for VO before biopsy and no associated treatment ( $n = 11$ ), and when clinical course was primarily influenced by comorbidities ( $n = 19$ ) such as other infection, malignancy, and/or cardiac arrest.

A final diagnosis of VO treated with antibiotics was documented in clinical notes of 70% (218/310) of cases, including 104 culture-positive cases and 114 presumed false-negative cultures. VO was excluded based on clinical factors and biopsy results in 20% (62/310) with alternate diagnoses including degenerative ( $n = 19$ ), trauma/fracture ( $n = 7$ ), inflammatory/rheumatologic ( $n = 6$ ), prior treated osteomyelitis ( $n = 4$ ), postradiation ( $n = 1$ ), and unspecified ( $n = 25$ ). The final diagnosis was unclear in 10% (30/310) of culture-negative or culture-contaminant cases in which treatment was also targeted to other pre-existing infections and comorbidities. The concordance rate of initial imaging impression and final clinical diagnosis was 63% (184/290). Management change was not significantly different between the imaging suspicious, equivocal, or not suspicious groups.

### DISCUSSION

Diagnosing VO is challenging given its insidious onset, nonspecific symptoms, overlapping clinical features with noninfectious etiologies and low biopsy yield.<sup>3–6</sup> Ideal management requires isolation of a causative organism through blood culture or biopsy followed by targeted antibiotic therapy.<sup>2</sup> Many patients are started on empiric antibiotics before biopsy. Given these factors,

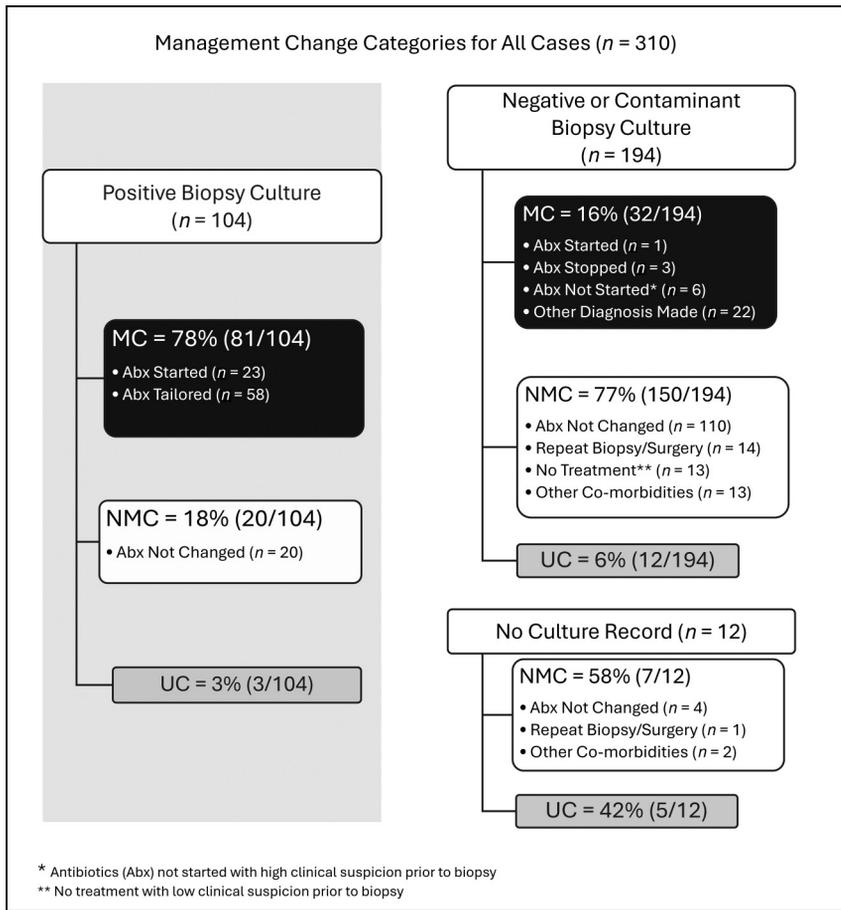


**FIG 2.** Isolated organisms on biopsy: true-positive (solid black), contaminant (striped).

we aimed to assess how image-guided biopsy results influence clinical management in suspected VO.

Similar to prior studies, we observed a suboptimal biopsy yield of 34% (104/310),<sup>3-6</sup> with *S aureus* being the most common

isolated pathogen.<sup>1,2</sup> We observed a modest overall change in management (36%, 113/310) due to the biopsy results, similar to the 37.5% results from Winkler et al<sup>8</sup> but variable from the results from Kuo et al (9%)<sup>10</sup> and Ang et al (50%).<sup>3</sup> Management was



**FIG 3.** Management change categories for all cases.

**Table 2: Clinical information relating to management change**

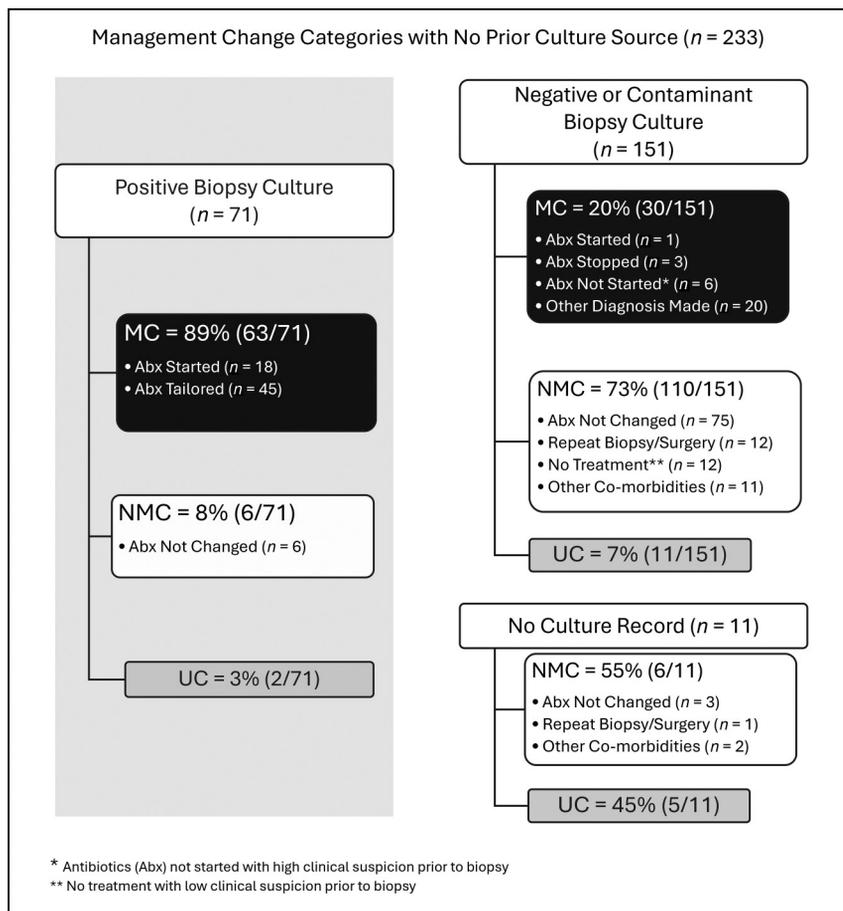
	Management Change (MC)	No Management Change (NMC)	Unclear	P Value (MC vs NMC)
<b>Biopsy</b>				
All biopsies	36% (113/310)	57% (177/310)	6% (20/310)	
Positive culture	72% (81/113)	11% (20/177)	3% (3/104)	<i>P</i> < .0001
Negative culture	27% (30/113)	79% (139/177)	6% (10/179)	<i>P</i> < .0001
Contaminant culture	2% (2/113)	9% (11/177)	13% (2/15)	<i>P</i> = .017
<b>Antibiotics</b>				
On antibiotics	19% (21/113)	40% (70/177)	3% (3/94)	<i>P</i> = .0002
No antibiotics	78% (88/113)	57% (101/177)	7% (15/204)	<i>P</i> = .0003
<b>Cultures</b>				
Other culture positive (any source)	18% (20/113)	31% (55/177)	3% (2/77)	<i>P</i> = .014
Blood culture positive	9% (10/113)	19% (33/177)	2% (1/44)	<i>P</i> = .021
No other culture source	82% (93/113)	69% (122/177)	8% (18/233)	<i>P</i> = .014
<b>Laboratory data</b>				
Abnormal WBC	13% (15/113)	23% (40/177)	0% (0/55)	<i>P</i> = .04
Elevated ESR and/or CRP	73% (82/113)	72% (128/177)	2% (5/215)	<i>P</i> = .85
<b>Imaging</b>				
Suspicious	75% (85/113)	75% (133/177)	6% (13/231)	<i>P</i> = 1.00
Equivocal	14% (16/113)	11% (20/177)	10% (4/40)	<i>P</i> = .45
Not suspicious	6% (7/113)	6% (11/177)	5% (1/19)	<i>P</i> = 1.00

impacted in most (78%, 81/104) culture-positive cases through tailoring of antibiotic regimens. This highlights the importance of optimizing biopsy methods for the greatest biopsy yield to impact management.

Even when an organism cannot be identified with biopsy, many patients are treated empirically if there is high clinical suspicion for VO as was the case in 57% (110/194) of the culture-negative or culture-contaminant biopsies. Despite 30% (94/310) of patients receiving antibiotics before biopsy, there was no significant difference in the biopsy yield based on short-term antibiotic exposure. Our results are in agreement with a recent meta-analysis by Chang et al<sup>5</sup> that reported no difference in microbiology sensitivities based on antibiotic exposure, though other studies report antibiotics influence culture yield.<sup>11</sup> Longer antibiotic duration (6.35 days) was associated with negative culture results compared with 3.69 days for positive culture results, though this difference did not reach statistical significance, suggesting that antibiotic status should not impact the decision to perform a biopsy but performing the biopsy sooner after initiating antibiotics may improve yield.

Higher biopsy yield (50%; 22/44) occurred in the setting of prior positive blood cultures. However, management change was significantly greater with no positive blood culture (MC 41%, 88/216) versus positive blood cultures (MC 23%, 10/44). This is likely because most patients with positive blood cultures (73%, 32/44) were on appropriate targeted antibiotics at biopsy with subsequent concordant or negative biopsy results that did not change therapy. Similarly, management change was greater in patients not on antibiotics at biopsy (MC 41%, 88/216) versus those on antibiotics (MC 22%, 21/94).

Fluid collection aspiration, disc biopsy (versus bone), withholding antibiotics, elevated CRP, and use of larger-bore needles have been associated with higher diagnostic biopsy yield,<sup>3,4,11,12</sup> though other studies question the importance of some of these factors.<sup>5</sup> We found disc sampling with or without bone (yield 36%, 82/228) and disc/bone



**FIG 4.** Management change categories with no prior culture source.

aspiration with or without cores (yield 42%, 39/92) were significantly associated with higher yield than bone only (yield 8%, 2/24) or core only (yield 29%, 56/193). The highest yield (56%, 10/18) was with aspiration only, possibly due to discrete fluid collections rather than solid material. However, variable reporting prevented us from determining if aspiration was from a discrete fluid collection or disc space in most cases. Needle size varied (9 g–22 g), though most used large-bore needles (14 g), with no difference between positive and negative culture groups, contrary to the conclusions from Hussein et al.<sup>12</sup>

Management change occurred in 16% (32/194) of negative/contaminant culture results or 10% (32/310) of all cases, similar to Winkler et al (11.7%).<sup>8</sup> Negative culture results influenced management when an alternate diagnosis was identified histologically. Core samples were crucial for identifying alternate diagnoses and ruling out infection.

With respect to image-guided biopsy for VO, the 2015 Infectious Disease Society of America (IDSA) guidelines<sup>2</sup> recommendations include the following: 1) defer image-guided biopsy when blood cultures are positive for *S aureus* or *S Lugdunensis* or testing is positive for *Brucella* species in endemic regions; 2) obtain image-guided biopsy when a microbiologic diagnosis is not established from blood cultures or serology; 3) obtain pathology specimens to guide therapy, which can be informative when cultures are

negative; and 4) withhold antimicrobial therapy before biopsy in stable patients to optimize culture sensitivity. These guidelines were not uniformly followed in this heterogeneous multisite retrospective study with a 25-year inclusion period, though our results support the first 3 listed recommendations. We found that short-term antibiotic administration did not impact biopsy yield, suggesting withholding antibiotics before emergent image-guided biopsy may not be necessary.

Biopsy yield is likely impacted by sample handling and laboratory processing. The IDSA recommends using anaerobic transport media and specimen delivery within 2 hours,<sup>2</sup> but these protocols could not be verified in our study. The amount of tissue available for microbiology can limit culture recovery, as samples are often shared with pathology. The variability of procedural techniques and sample processing practices in this study is a limitation, but we believe our multi-center design without shared technical policies is reflective of general hospital practices, increasing the external validity of results. The inclusion of cases with variable imaging interpretations, procedural modalities, and target tissue types provides useful information, which may have served as exclusion criteria in prior studies.<sup>8,10,11</sup>

This study is limited by the subjective nature of deciding management change retrospectively, though we used objective measures related to antibiotic therapy to determine management change. Additionally, we were limited by variable reporting of biopsy technique, including tissue sampled, needle size, and sample number. Laboratory methods were not standardized across time and sites, which could impact biopsy yield, and we could not determine which exact sample led to a positive diagnosis. These factors have been examined in prior studies and emphasize the benefit of standardized procedure reporting to optimize our understanding of impactful procedural practices. Future prospective studies could be performed to contemporaneously evaluate biopsy techniques, laboratory methods, and clinical scenarios resulting in the greatest management changes.

This study is limited by the subjective nature of deciding management change retrospectively, though we used objective measures related to antibiotic therapy to determine management change. Additionally, we were limited by variable reporting of biopsy technique, including tissue sampled, needle size, and sample number. Laboratory methods were not standardized across time and sites, which could impact biopsy yield, and we could not determine which exact sample led to a positive diagnosis. These factors have been examined in prior studies and emphasize the benefit of standardized procedure reporting to optimize our understanding of impactful procedural practices. Future prospective studies could be performed to contemporaneously evaluate biopsy techniques, laboratory methods, and clinical scenarios resulting in the greatest management changes.

## CONCLUSIONS

This study assessed how biopsy results influence the clinical management of VO. Management change is greatest with positive biopsy culture results, no prior culture sources, and among patients not on antibiotics at biopsy. These results support performing a biopsy when a causative organism has not been isolated from another source. Antibiotic therapy at the time of biopsy does not impact yield but may lead to less management change. Disc

sampling and aspiration are associated with higher culture yield. Negative culture results also influence management, and core biopsy identifies alternate diagnoses.

## ACKNOWLEDGMENTS

We thank Thomas E. Grys, PhD, Associate Professor and Co-Director of Microbiology at the Mayo Clinic Arizona, for his contribution to the methods and discussion.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Gouliouris T, Aliyu SH, Brown NM. **Spondylodiscitis: update on diagnosis and management.** *J Antimicrob Chemother* 2010;65 Suppl 3: iii11–iii24 CrossRef Medline
2. Berbari EF, Kanj SS, Kowalski TJ, et al. **2015 Infectious Diseases Society of America (IDSA) clinical practice guidelines for the diagnosis and treatment of native vertebral osteomyelitis in adults.** *Clin Infect Dis* 2015;61:e26–46–e46 CrossRef Medline
3. Ang MT, Wong GR, Wong DR, et al. **Diagnostic yield of computed tomography-guided biopsy and aspiration for vertebral osteomyelitis.** *J Med Imaging Radiat Oncol* 2019;63:589–95 CrossRef Medline
4. Sehn JK, Gilula LA. **Percutaneous needle biopsy in diagnosis and identification of causative organisms in cases of suspected vertebral osteomyelitis.** *Eur J Radiology* 2012;81:940–46 CrossRef Medline
5. Chang CY, Pelzl C, Jesse MK, et al. **Image-guided biopsy in acute diskitis-osteomyelitis: a systematic review and meta-analysis.** *AJR Am J Roentgenol* 2023;220:499–511 CrossRef Medline
6. Pupaibool J, Vasoo S, Erwin PJ, et al. **The utility of image-guided percutaneous needle aspiration biopsy for the diagnosis of spontaneous vertebral osteomyelitis: a systematic review and meta-analysis.** *Spine J* 2015;15:122–31 CrossRef Medline
7. Hoang D, Fisher S, Oz OK, et al. **Percutaneous CT guided bone biopsy for suspected osteomyelitis: diagnostic yield and impact on patient's treatment change and recovery.** *Eur J Radiology* 2019;114:85–91 CrossRef Medline
8. Winkler WL, George IA, Gandra S, et al. **Diagnostic efficacy and clinical impact of image-guided core needle biopsy of suspected vertebral osteomyelitis.** *Int J Infect Dis* 2024;144:107027 CrossRef Medline
9. Rankine JJ, Barron DA, Robinson P, et al. **Therapeutic impact of percutaneous spinal biopsy in spinal infection.** *Postgrad Med J* 2004;80:607–09 CrossRef Medline
10. Kuo AH, Cho CH, Huang RY, et al. **Target-specific yield rate and clinical utility of percutaneous tissue sampling in spinal infection.** *Clin Imaging* 2020;68:257–62 CrossRef Medline
11. Maamari JB, Tande AJ, Tai DBG, et al. **Factors impacting the yield of image-guided biopsy in native vertebral osteomyelitis: a 10-year retrospective study.** *Open Forum Infect Dis* 2022;9:ofac616 CrossRef Medline
12. Husseini JS, Simeone FJ, Nelson SB, et al. **CT-guided discitis-osteomyelitis biopsies: needle gauge and microbiology results.** *Skeletal Radiology* 2020;49:1431–39 CrossRef Medline

# Prevalence and Characteristics of Microspurs in Patients with Spontaneous Intracranial Hypotension Compared with the General Population

AQ:au  Danial Nasiri,  Levin Häni,  Johannes Goldberg,  Thomas Petutschnigg,  Tomas Dobrocky,  Ralph T. Schär,  Christoph Schankin,  Andreas Raabe,  Jürgen Beck,  Eike Immo Piechowiak, and  Christopher Marvin Jesse



## ABSTRACT

**BACKGROUND AND PURPOSE:** In patients diagnosed with spontaneous intracranial hypotension (SIH), microspurs are considered the culprit lesion in most ventral dural leaks (type I). The imaging characteristics of discogenic spurs, and their prevalence in the general population has not been reported in the literature.

**MATERIALS AND METHODS:** This observational case-control study was conducted comparing the prevalence and characteristics of discogenic microspurs between patients with SIH with a type I leak treated at a tertiary hospital between 2013 and 2023 and an age- and sex-matched cohort of trauma patients.

**RESULTS:** Each group consisted of 85 patients (mean age  $51.6 \pm 11.9$  years), 74% (58/85 patients) were women. The prevalence of discogenic microspurs in the control group and SIH group was 31.8% and 90.6%, respectively. The mean length of the culprit microspur responsible for a dural leak was larger compared with the mean length of all coincidental microspurs from both the SIH and the control group not causing a dural leak (2.6 mm versus 1.6 mm,  $P < .001$ ). Our multivariate logistic regression revealed that an increasing length of a microspur (OR, 1.942, CI 1.35–2.80,  $P < .001$ ) and a narrower diameter of the spinal canal (OR, 0.85, CI 0.76–0.96,  $P = .008$ ) were predictive for a dural tear.

**CONCLUSIONS:** A discogenic microspur is a common incidental finding and may be found in almost one-third of the general population. The length of the culprit microspur and the diameter of the spinal canal are distinct morphologic characteristics for type I associated CSF leaks.

AQ:abbr **ABBREVIATIONS:** SLEC = spinal longitudinal extradural fluid; SIH = spontaneous intracranial hypotension

Spontaneous intracranial hypotension (SIH) is a debilitating disease with an estimated incidence of 5 in 100,000 cases per year.<sup>1–9</sup> Spinal CSF leaks are divided into 3 types: ventral leaks (type I), lateral leaks (type II), and CSF-venous fistula (type III).<sup>10</sup>

SIH is primarily caused by a CSF leak in the spinal canal where in type I lesions a microspur can often be found at the leakage site.<sup>11–13</sup> These microspurs are thought to be calcifications of the intervertebral disc usually measuring just a few millimeters.<sup>14</sup>

In type I leaks, most microspurs are located at the cervicothoracic and thoracolumbar junction.<sup>8,9,14–17</sup> The prevalence and location of microspurs in the general population has not been reported. It is further unclear why patients with SIH with multiple microspurs show only 1 leakage site whereas the other microspurs are not causing a dural tear. So far, there are no studies investigating the prevalence of microspurs in the general population and analyzing the characteristics of microspurs that lead to a dural tear and thereby SIH.

The aim of our study was to compare the prevalence and imaging characteristics of microspurs in the general population and patients with SIH type I. We hypothesized that microspur findings in the general population are prevalent and show distinct morphologic characteristics in patients with SIH.

## MATERIALS AND METHODS

### Study Design

We conducted a retrospective, observational case-control study. CT myelographies of patients with SIH type I with proved ventral CSF leaks from our SIH database were analyzed for the

Received November 19, 2024; accepted after revision December 30.

From the Departments of Neurosurgery (D.N., L.H., J.G., T.P., R.T.S., A.R., C.M.J.), and Neurology (C.S.), and Institute of Diagnostic and Interventional Neuroradiology (T.D., E.I.P.), Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland; and Department of Neurosurgery (J.B.), Medical Center–University of Freiburg, Freiburg, Germany.

Eike Immo Piechowiak and Christopher Marvin Jesse contributed equally to this work.

AQ:corr Please address correspondence to Danial Nasiri, MD, Department of Neurosurgery, Inselspital, Bern University Hospital, Bern, Switzerland; e-mail: danial.nasiri@insel.ch; @DanialNasiri94



Indicates article with supplemental data.

AQ:perm <http://dx.doi.org/10.3174/ajnr.A8644>

## SUMMARY

**PREVIOUS LITERATURE:** Previous studies on SIH type I have provided insights into the nature of spicules (osteogenic versus fibrous) that contribute to dural tears, as well as the baseline characteristics of patients predisposed to developing a ventral dural leak. These patients are typically described as tall, slender, and women. Additionally, numerous case reports highlight the often protracted diagnostic journey for SIH type I and discuss various surgical techniques for treating these patients. It is also frequently noted that the disease tends to go undiagnosed due to its rarity.

**KEY FINDINGS:** We demonstrated that longer microspurs and narrower spinal canals increase the likelihood of a spinal dural leak. Furthermore, we found that 91% of patients with SIH type I have microspurs, while nearly 32% of the general population also exhibit microspurs.

**KNOWLEDGE ADVANCEMENT:** The high prevalence of microspurs in the general population is a novel finding, suggesting that not all microspurs are responsible for causing a dural tear. Additionally, this is the first study to describe the characteristics of microspurs that support the intuitive assumption that larger microspurs and a narrow spinal canal are key factors in the development of a dural tear.

parameters explained in a later section. This was equally performed for trauma patients without spinal findings who received a whole-spine CT at our institution. Approval from the local ethics committee of the canton of Bern, Switzerland, was obtained (2020–00645). The Strengthening the Reporting of Observational Studies in Epidemiology checklist was used as a validation tool and checklist for our case-control study.

AQ: A

### Patient Selection and Data Collection

We included patients with SIH with a ventral CSF leak (type I),<sup>10</sup> treated at our institution from February 2013 to October 2023, and compared them with a control group consisting of patients who entered the emergency department and received a whole-spine CT from March 2018 until February 2024.

Patients in the control cohort were matched based on sex and age without considering their primary disease or reason for admission. The patients in the control group were selected randomly to minimize selection bias. We then analyzed each patient in the SIH and control cohort and screened the dynamic CT myelography or CT scans for microspurs, respectively. In most cases, an unenhanced CT scan was performed during dynamic CT myelography or for preoperative planning. Only ventrally located microspurs were included. Exclusion criteria were congenital deformities of the spine, spinal injuries on imaging, SIH type II and III leaks as well as low image quality that did not allow for accurate diagnostic measurements of a microspur.

Additionally, we compared 2 subgroups of microspurs: the culprit microspurs associated with a dural leak (only SIH cohort) and all coincidental microspurs of SIH and control patients without an associated dural leak.

The length and width of each microspur was measured on transversal images at the base of the microspur (Fig 1, measurements b and c). Next, we measured the diameter of the spinal canal at the site of the culprit and/or coincidental microspurs (Fig 1, measurement a). Finally, measurements for the deviation of the microspur from the midsagittal plane were performed (Fig 1, measurement d). The laterality of the microspur was defined as the offset from the midsagittal plane and was measured on the transversal plane (in millimeters). All measurements were strictly performed on the level of the intervertebral disc on multiplanar

reconstruction to exclude spiculae at the level of the vertebral body. For our imaging analysis, we utilized the SECTRA IDS7 software. Figure 1 illustrates how each parameter was measured.

### Classification of SIH Leaks

According to Schievink et al,<sup>10</sup> 3 types of SIH leaks can be identified. Type I refers to a dural tear that can be divided into type Ia, ventral leaks, and type Ib, posterolateral CSF leaks. These SIH leaks are often associated with a spinal longitudinal extradural fluid (SLEC) collection. Type II consists of meningeal diverticula that less often result in an extradural CSF collection. Last, type III leaks comprise CSF-venous fistula that are not associated with an extradural CSF collection.

### Statistics

Descriptive data included calculation of the mean, standard deviation (SD), and confidence interval (CI). Normal distribution was tested for all parameters considered. A Mann-Whitney *U* test was conducted for comparison of continuous parameters. Both univariate and multivariate binary logistic regression was performed for testing of our 3 microspur parameters mentioned above and the baseline characteristics (length and width of microspurs, diameter of the spinal canal, age, sex). Statistical significance of variables was considered for a *P* value less than .05.  $\chi^2$  testing was performed for categorical variables. Statistical analysis was performed by using the statistical software SPSS (Version 28.0.1.1, IBM).<sup>14</sup>

## RESULTS

### Patient Characteristics and Prevalence of Microspurs in SIH and Control Groups

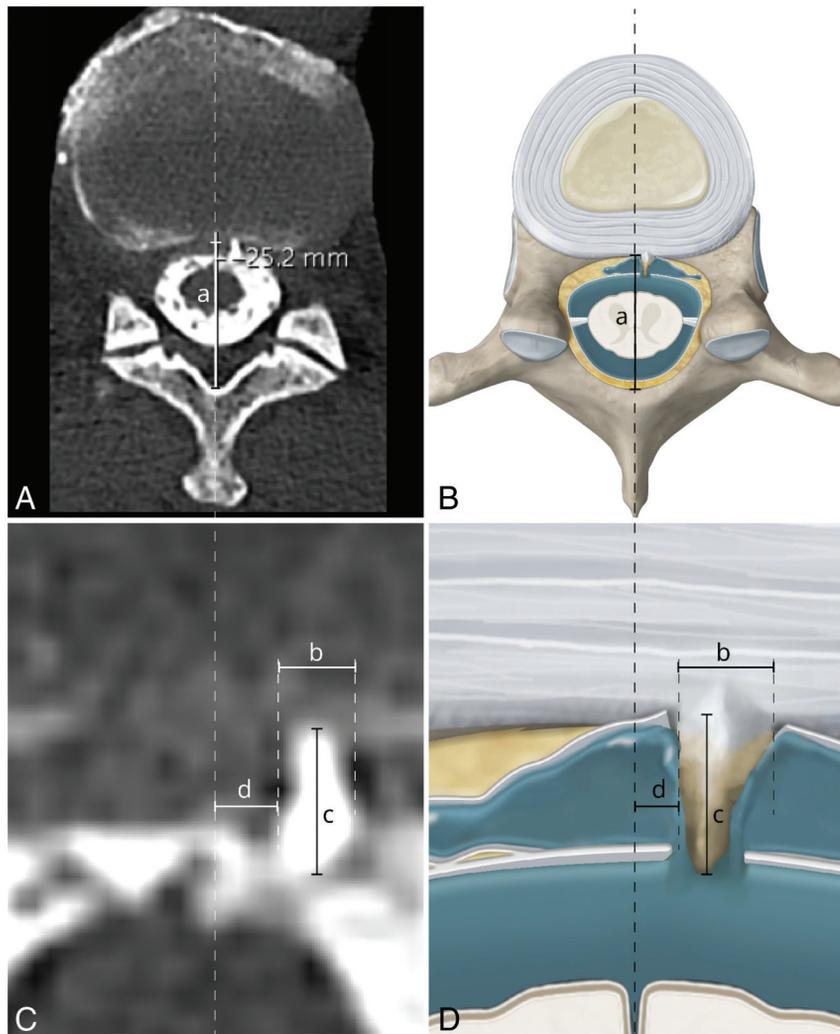
In total, 85 patients with SIH demonstrated a type I leak during the study period. In 77 of them (91%), the spinal CSF leak was due to a microspur. In the control group, 27 of 85 (32%) showed at least 1 microspur.

Mean age was 51.8 years (SD  $\pm$  11.59 years, 95% CI 49.3–54.3 years) and 58 patients (68%) were women (Table 1).

T1

### Number of Microspurs between SIH and Control Groups

The number of microspurs per person ranged from zero to 6 in both groups. The mean number of microspurs among the SIH



**FIG 1.** The top row displays the measurement of the spinal canal diameter (A). The bottom row represents the measurements of the microspur's width (B), length (C), and its deviation from the midline (D).

**Table 1: Baseline characteristics of patients**

	n	SIH	Control
Age (+ SD 11.59)	85	51.8	51.8
Female	85	0.68	0.68

**Table 2: Number of microspurs between the SIH and control groups**

	Control	SIH	Total
Number of microspurs			
0	58	8	66
1	13	45	58
2	7	15	22
3	3	8	11
4	1	4	5
5	3	4	7
6	0	1	1
Total	85	85	170

**Note:**—The microspurs were detected at different levels along the spinal canal. At each level, a single microspur was found.

group was 1.7 (SD ± 1.3, 95% CI 1.4–1.9) and 0.6 (SD ± 1.2, 95% CI 0.4–0.9) in the control group, showing a statistically significant difference ( $P < .001$ ).

For both groups, considering only those with microspurs on imaging, most showed 1 microspur (58% in the SIH group, 48% in the control group) (Table 2).

**Location of Culprit and Coincidental Lesions along the Spinal Axis**

The most frequent location of culprit microspurs causing a dural leak was at T1-T2 (17%) and T11-T12 (15%). For co-incidental microspurs, the most frequent location was at T2-T3 (11%) and T1-T2 and T10-T11 with equally 10% (Fig 2).

**Location of Culprit Microspurs Causing a Leak (Midline versus Lateral)**

The comparison of culprit and coincidental microspurs relative to the midline of the spinal canal (midline versus lateral) showed that culprit lesions were more likely to be centrally located ( $P = .01$ ) (Table 3).

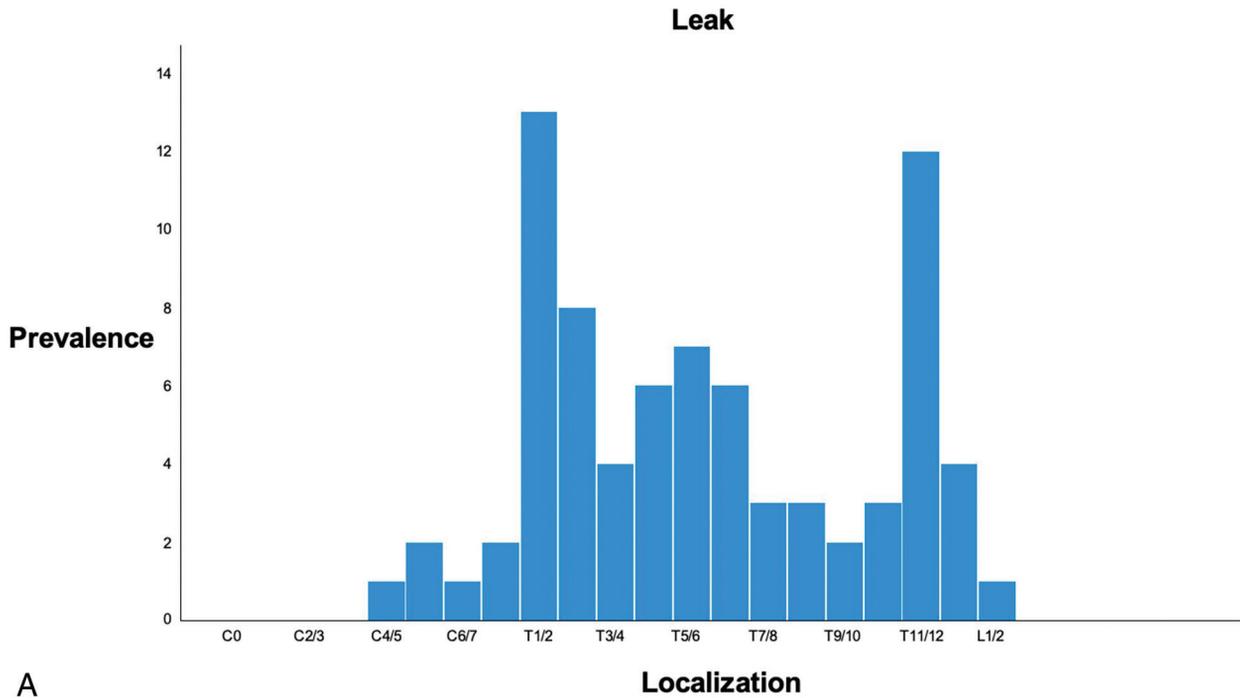
**Microspur and Spinal Anatomic Parameters**

The mean length of a microspur that caused a ventral dural leak (2.6 mm) was larger compared with those that did not cause a dural leak (1.6 mm) (2.6 mm versus 1.6 mm, SD = 1.48,  $P < .001$ ).

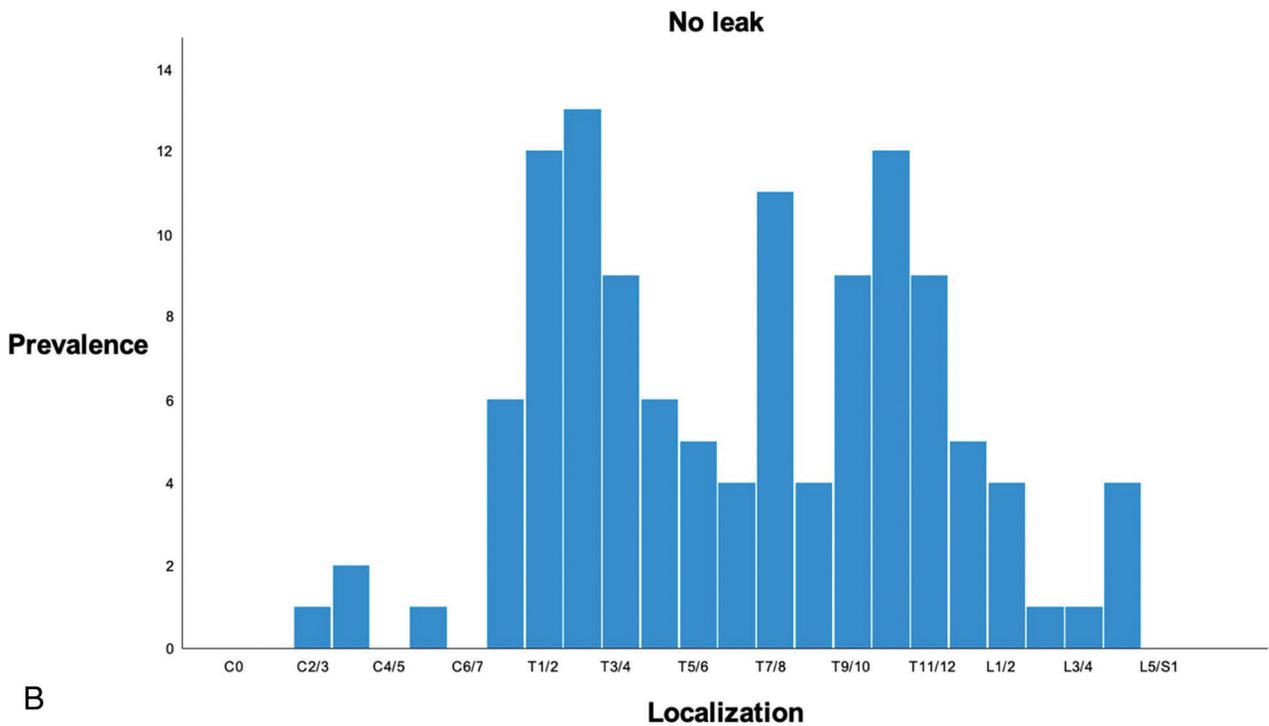
The univariate analysis revealed that the length (OR, 1.68, 95% CI 1.29–2.20;  $P < .001$ ) and width (OR, 1.26, CI 1.05–1.52;  $P = .01$ ) of a microspur as well as the diameter of the spinal canal (OR, 0.89,

CI 0.80–0.99;  $P = .03$ ) show a statistically significant difference between the culprit lesion group (associated dural leak) compared with the coincidental microspur cohort without a dural leak. Therefore, our univariate analysis revealed that a longer and wider microspur as well as a narrower spinal canal are causal for a dural leak.

In our multivariate analysis, these results were only robust for 2 of the former parameters. The length of the microspur (OR, 1.942, CI 1.35–2.80,  $P < .001$ ) was still indicated to be a significant factor in spinal leakage, with longer microspurs being more likely to cause a tear in the dura. Furthermore, a smaller diameter of the spinal canal (OR, 0.85, CI 0.76–0.96,  $P = .008$ ) is a risk factor for a dural slit. However, the width of the microspur (OR, 0.93, CI 0.72–1.19,  $P = .545$ ) showed no statistically significant difference between our culprit lesion group and coincidental group in the multivariate analysis. There was no statistically significant difference in the diameter of the spinal canal at the microspur sites between men and women (20.46 mm versus 19.76 mm, SD = 2.78,  $P = .242$ ). Table 4 illustrates our findings of the univariate and multivariate analysis.



A



B

**FIG 2.** A, The distribution of culprit microspsurs responsible for dural leaks is shown. B, The distribution of coincidental microspsurs that do not lead to a dural leak is illustrated. The spinal level is represented on the x-axis, with the prevalence for each level indicated on the y-axis.

**Table 3: Position of microspsurs and association with dural leak**

Side	Total	Coincidental Spsurs	Culprit Spsurs
Center	54	25	29
Lateral	143	94	49
Total	197	119	78

**Note:**—Center refers to a ventral, strictly midline located microspsur. Lateral refers to an either left-sided or right-sided microspsur in reference to the midline.

**DISCUSSION**

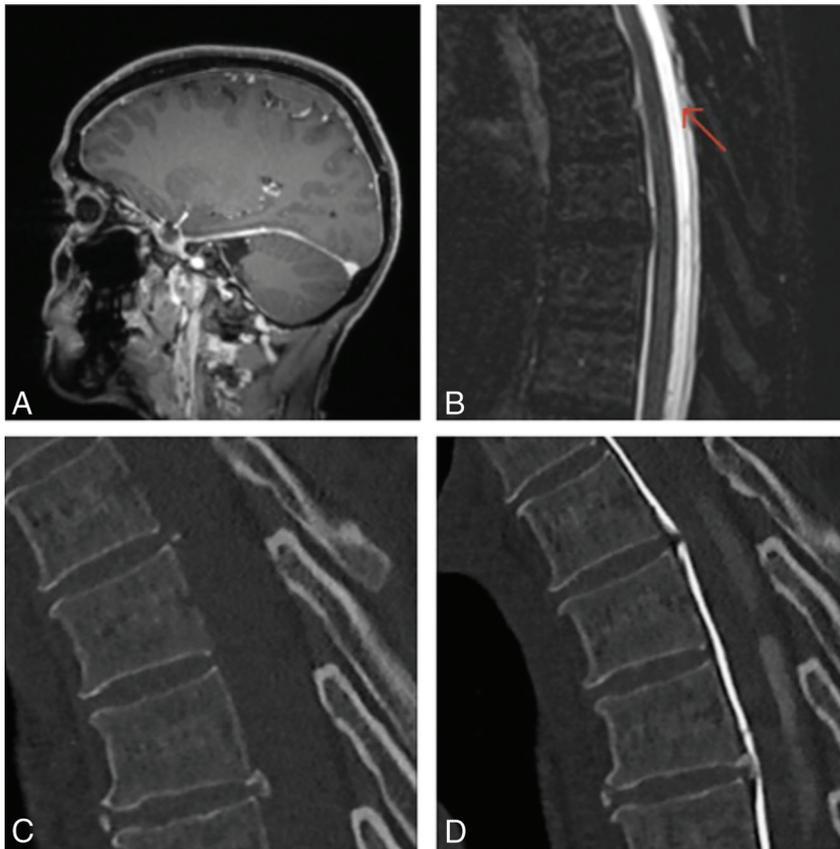
This study confirms that microspsurs are prevalent in the general population. Our findings indicate that both the length of microspsurs and the diameter of the spinal canal are significant factors contributing to dural tears, which are implicated in type I associated SIH (Fig 3).

F3

**AUTHOR: PLEASE NOTE GRAPHIC QUALITY**

**Table 4: Univariate and multivariate analysis of predictive factors for dural leaks due to microspurs**

Variable	Univariate		Multivariate	
	P Value	OR (95% CI)	P Value	OR (95% CI)
Length	<.001	1.683 (1.287–2.200)	<.001	1.942 (1.347–2.800)
Width	.013	1.263 (1.050–1.519)	.545	0.926 (0.720–1.189)
∅ of spinal canal	.034	0.889 (0.798–0.991)	.008	0.851 (0.756–0.958)



**FIG 3.** This is a case of a 66-year-old patient presenting with an orthostatic headache. A, Typical MRI findings suggestive of SIH. B, Longitudinal extradural fluid along the spine is visible (indicated by the red arrow). C and D, Native and contrast-enhanced CT scans from dynamic CT myelography are shown, revealing a microspur responsible for a dural tear.

We hypothesized that the length of a microspur is a critical determinant for the occurrence of a dural tear. Moreover, the narrowing of the spinal canal, particularly prominent in the thoracic spine, appears to increase the risk of a dural leak. This could be due to the restricted space for microspurs to stay clear of the dura along with the shift from lordosis to kyphosis and back again at the junction sites. Consistent with previous studies, we found that almost all microspurs in both SIH and non-SIH patients were located in the thoracic region, predominantly at the cervicothoracic and thoracolumbar junctions.<sup>8,9,14-17</sup> The rigidity and reduced diameter of the thoracic spine may predispose it to dural leaks compared with the cervical or lumbar regions.

In our control group, 32% exhibited microspurs, while 91% of our SIH cohort with ventral type I leaks had microspurs. The high prevalence of microspurs in the general population raises the

question of how many patients go on to develop SIH. It further leaves unanswered why some patients never develop SIH despite this anatomic feature. It is well-established that SIH is often underdiagnosed, partly due to nonspecific symptoms, particularly orthostatic headache, which may go unrecognized at the onset.<sup>8,13</sup>

Beck et al<sup>14</sup> noted that the mere presence of a microspur does not necessarily correlate with a dural CSF leak. To investigate the causes of dural slits in the spinal meninges further, we categorized our subjects into 2 groups: those with a dural leak and those without. This classification enabled us to analyze microspur characteristics associated with dural leaks and understand why patients with SIH type I with multiple microspurs exhibit only 1 leakage site. Our analyses demonstrated that an increased microspur length and a reduced spinal canal diameter significantly increased the probability of a dural leak. Rosebrock et al<sup>18</sup> have demonstrated that patients with type I leaks often have a lower body mass index, which increases their susceptibility to developing a spinal dural leak due to reduced epidural fat. In our study, we concentrated exclusively on the osseous boundaries of the spinal canal, without addressing the epidural fat.

The nonsignificant finding regarding microspur width in the multivariate analysis is understandable; wider microspurs tend to extend laterally into the subarachnoid space rather than toward the dura surrounding the spinal cord. The significant finding for width in the

univariate analysis likely results from the correlation between broader microspurs and overall size.

Our demographic data align with existing literature on patients with SIH, predominantly middle-aged women.<sup>9,16,19</sup> Although the length of microspurs may not differ by sex, we explored whether the spinal canal diameter in women is smaller than in men at the microspur sites, potentially explaining the higher prevalence of SIH in women. However, we found no statistically significant correlation between sex and SIH development. The increased prevalence in women may be attributed to earlier symptom presentation<sup>20</sup> and the higher incidence of headaches, leading to more frequent medical consultations.<sup>21,22</sup>

Distinguishing pathogenic lesions in SIH is complex, because they also occur in the general population. This complexity applies to both type I leaks, as well as type II leaks. A study by Kranz et al<sup>23</sup> found a positive trend without statistical significance in the

prevalence and myelographic appearance of meningeal diverticula between patients with SIH and controls, despite known associations. The presence and quantity of microspurs increase the risk for dural tears as more locations along the spinal axis are vulnerable to a dural tear.<sup>14,24</sup> While we identified differences in microspur prevalence and characteristics between type I leaks and controls, our data show that most patients present a single microspur, which raises questions about the relationship between microspur quantity and dural leaks. Our analysis revealed that many patients in both groups had only 1 microspur; however, 42% of the SIH group exhibited multiple microspurs. These results are consistent with previous findings from Beck et al,<sup>14</sup> underscoring the importance of accurately locating the dural tear, particularly in patients with multiple microspurs, because this is critical for effective treatment. Another important consideration is the rare case in which some microspurs tend to vanish over time.<sup>25</sup> This might also explain the few SIH cases in our cohort that do not show a microspur at the leakage site, which can prolong diagnosis and treatment of SIH.

We also examined whether microspurs associated with dural leaks were more likely to be found medially versus laterally. Our findings indicate a significant correlation between centrally located microspurs and the likelihood of causing a dural tear. This is in agreement with other studies that mainly found microspurs to be found in the midline.<sup>26,27</sup> However, there are also small series of lateral dural tears associated with bony spicules.<sup>28</sup> Despite this rare finding, our results could explain the increased likelihood of a dural tear for lateral bony spicules.

In his recent publication, Takai et al<sup>26</sup> discusses the etiology of a Hofmann ligament as the cause for a dural slit rather than calcified disc herniations, despite midline findings of microspurs. The Hofmann ligament is considered an epidural ligament arising from the posterior longitudinal ligament.<sup>29,30</sup> In their study, Takai et al<sup>26</sup> report to have found soft and easily removable microspurs that upon histologic analysis were revealed to be fibrotic tissue without osteophytic or degenerative calcified components. This is interesting as most of the literature on SIH considers discogenic microspurs to be the main cause for dural slits. The findings from the Tokyo group<sup>26</sup> in Japan may justify histologic analysis of microspur removal to determine whether microspurs originate from calcified intervertebral discs or if they arise from the Hofmann ligament.

Understanding specific characteristics of microspurs that may increase the likelihood of causing a dural tear in SLEC-positive patients would be highly beneficial in the diagnostic evaluation. This knowledge could enable more targeted planning of myelographic studies, thereby optimizing the diagnostic process and potentially reducing the radiation exposure, which is notably high in the work-up for SIH.<sup>31</sup>

Our study has several limitations. First, it is a retrospective analysis that did not account for patient comorbidities, such as collagenoses, osteoporosis, prior disc herniations, or other conditions that may contribute to the development of discogenic microspurs. Second, our control group was defined based on sex, age, and imaging quality, without considering other patient-related factors. However, this approach may have reduced selection bias. Another limitation is the presence of blooming artifacts due to the attenuated

layer of contrast agent in close proximity to the microspur, which can lead to measurement errors. Last, we cannot rule out the possibility that any of our trauma patients may have had SIH.

## CONCLUSIONS

While osteodiscogenic microspurs occur regularly in patients with type I associated CSF leaks, they are also a common incidental finding in the general population. Anatomic features, such as the length of microspurs and the diameter of the spinal canal, are distinctive characteristics to microspurs associated with a CSF leak. The close proximity of microspurs to the spinal canal, identified as the primary risk factor for a dural slit, highlights the mechanical nature of SIH type I leaks.

Disclosure forms provided by the authors are available with the full text and AQ:disc PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Lin J, Ping Zhang S, Dong He F, et al. **The status of diagnosis and treatment to intracranial hypotension, including SIH.** *J Headache Pain* 2017;18:4 [CrossRef Medline](#)
2. Wolf K, Volz F, Lützen N, et al. **Non-invasive biomarkers for spontaneous intracranial hypotension (SIH) through phase-contrast MRI.** *J Neurol* 2024;271:4336–47 [CrossRef Medline](#)
3. Jesse CM, Häni L, Fung C, et al. **The impact of spontaneous intracranial hypotension on social life and health-related quality of life.** *J Neurol* 2022;269:5466–73 [CrossRef Medline](#)
4. Brosnan C, Dablouk MO, Healy V, et al. **When does too little pressure become too much? A case of spontaneous intracranial hypotension presenting with acute loss of consciousness.** *Surg Neurol Int* 2023;14:338 [CrossRef Medline](#)
5. Cheema S, Anderson J, Angus-Leppan H, et al. **Multidisciplinary consensus guideline for the diagnosis and management of spontaneous intracranial hypotension.** *J Neurol Neurosurg Psychiatry* 2023;94:835–43 [CrossRef Medline](#)
6. Jesse CM, Schär RT, Petutschnigg T, et al. **Improvement of health-related quality of life after closure of spinal CSF leaks in patients with spontaneous intracranial hypotension.** *J Neurosurg Spine* 2024;41:452–58
7. Mokri B. **Spontaneous low pressure, low CSF volume headaches: spontaneous CSF leaks.** *Headache J Head Face Pain* 2013;53:1034–53 [CrossRef Medline](#)
8. Schievink WI. **Spontaneous spinal cerebrospinal fluid leaks and intracranial hypotension.** *JAMA* 2006;295:2286–96 [CrossRef Medline](#)
9. Schievink WI, Maya MM, Moser F, et al. **Frequency of spontaneous intracranial hypotension in the emergency department.** *J Headache Pain* 2007;8:325–28 [CrossRef Medline](#)
10. Schievink WI, Maya MM, Jean-Pierre S, et al. **A classification system of spontaneous spinal CSF leaks.** *Neurology* 2016;87:673–79 [CrossRef Medline](#)
11. Goldberg J, Häni L, Jesse CM, et al. **Spontaneous intracranial hypotension without CSF leakage—concept of a pathological cranial to spinal fluid shift.** *Front Neurol* 2021;12:760081 [CrossRef Medline](#)
12. Schievink WI, Maya M, Moser F, et al. **Long-term risks of persistent ventral spinal CSF leaks in SIH: superficial siderosis and bibrachial amyotrophy.** *Neurology* 2021;97:e1964–70 [CrossRef Medline](#)
13. Beck J, Fung C, Ulrich CT, et al. **Cerebrospinal fluid outflow resistance as a diagnostic marker of spontaneous cerebrospinal fluid leakage.** *J Neurosurg Spine SPI* 2017;27:227–34 [CrossRef Medline](#)
14. Beck J, Ulrich CT, Fung C, et al. **Diskogenic microspurs as a major cause of intractable spontaneous intracranial hypotension.** *Neurology* 2016;87:1220–26 [CrossRef Medline](#)

15. Matsuhashi A, Takai K, Taniguchi M. **Microsurgical anatomy and treatment of dural defects in spontaneous spinal cerebrospinal fluid leaks.** *J Neurosurg Spine* 2021;34:522–30 [CrossRef Medline](#)
16. Beck J, Häni L, Ulrich CT, et al. **Diagnostic challenges and therapeutic possibilities in spontaneous intracranial hypotension.** *Clin Transl Neurosci* 2018;2:2514183X1878737 [CrossRef](#)
17. Dobrocky T, Nicholson P, Häni L, et al. **Spontaneous intracranial hypotension: searching for the CSF leak.** *Lancet Neurol* 2022;21:369–80 [CrossRef Medline](#)
18. Rosebrock RE, Diehn FE, Luetmer PH, et al. **Penetrating osseous spicules causing high-flow ventral CSF leaks in the setting of relatively low BMI: a preliminary study.** *Clin Neuroradiol* 2018;28:539–43 [CrossRef Medline](#)
19. Piechowiak EI, Pospieszny K, Haeni L, et al. **Role of conventional dynamic myelography for detection of high-flow cerebrospinal fluid leaks: optimizing the technique.** *Clin Neuroradiol* 2021;31:633–41 [CrossRef Medline](#)
20. Lin P-T, Hseu S-S, Fuh J-L, et al. **Sex differences in the clinical manifestations and treatment outcomes in a large cohort of spontaneous intracranial hypotension.** *Headache J Head Face Pain* 2024;64:1298–308 [CrossRef Medline](#)
21. Victor T, Hu X, Campbell J, et al. **Migraine prevalence by age and sex in the United States: a life-span study.** *Cephalalgia* 2010;30:1065–72 [CrossRef Medline](#)
22. Al-Hassany L, Haas J, Piccininni M, et al. **Giving researchers a headache—sex and gender differences in migraine.** *Front Neurol* 2020;11:549038 [CrossRef Medline](#)
23. Kranz PG, Stinnett SS, Huang KT, et al. **Spinal meningeal diverticula in spontaneous intracranial hypotension: analysis of prevalence and myelographic appearance.** *AJNR Am J Neuroradiol* 2013;34:1284–89 [CrossRef Medline](#)
24. Schievink WI, Maya MM, Moser F, et al. **Multiple spinal CSF leaks in spontaneous intracranial hypotension: do they exist?** *Neurol Clin Pract* 2021;11:e691–97 [CrossRef Medline](#)
25. Schievink WI, Ross L, Prasad RS, et al. **Vanishing calcification associated with a spontaneous ventral spinal cerebrospinal fluid leak.** *Cephalalgia* 2016;36:1366–69 [CrossRef Medline](#)
26. Takai K, Endo T, Komori T. **Calcified Hofmann’s ligaments as the cause of spinal cerebrospinal fluid leaks associated with spinal ventral dural tears.** *J Neurosurg Spine* 2025;42:43–48 [CrossRef Medline](#)
27. Häni L, Fung C, El Rahal A, et al. **Distinct pattern of membrane formation with spinal cerebrospinal fluid leaks in spontaneous intracranial hypotension.** *Oper Neurosurg (Hagerstown)* 2024;26:71–77 [CrossRef Medline](#)
28. Madhavan AA, Shlapak DP, Benson JC, et al. **Osseous spicules of the posterior elements causing fast cerebrospinal fluid leaks.** *Neuroradiology* 2022;64:1689–93 [CrossRef Medline](#)
29. Wiltse LL, Fonseca AS, Amster J, et al. **Relationship of the dura, Hofmann’s ligaments, Batson’s plexus, and a fibrovascular membrane lying on the posterior surface of the vertebral bodies and attaching to the deep layer of the posterior longitudinal ligament: an anatomical, radiologic, and clinical study.** *Spine (Phila Pa 1976)* 1993;18:1030–43 [CrossRef Medline](#)
30. Tardieu GC, Fisahn C, Loukas M, et al. **The epidural ligaments (of Hofmann): a comprehensive review of the literature.** *Cureus* 2016;8:e779 [CrossRef](#)
31. Dobrocky T, Mosimann PJ, Zibold F, et al. **Cryptogenic cerebrospinal fluid leaks in spontaneous intracranial hypotension: role of dynamic CT myelography.** *Radiology* 2018;289:766–72 [CrossRef Medline](#)

# Fusion of FDG and FMZ PET Reduces False-Positives in Predicting Epileptogenic Zone

Bingyang Cai, Shize Jiang, Hui Huang,  Jiwei Li, Siyu Yuan, Ya Cui, Weiqi Bao, Jie Hu, Jie Luo, and Liang Chen

## ABSTRACT

**BACKGROUND AND PURPOSE:** Epilepsy, a globally prevalent neurologic disorder, necessitates precise identification of the epileptogenic zone (EZ) for effective surgical management. While the individual utilities of FDG-PET and flumazenil (FMZ)-PET have been demonstrated, their combined efficacy in localizing the epileptogenic zone remains underexplored. We aim to improve the noninvasive prediction of EZ in temporal lobe epilepsy (TLE) by combining FDG-PET and FMZ-PET with statistical feature extraction and machine learning.

**MATERIALS AND METHODS:** This study included 20 drug-resistant patients with unilateral TLE (14 mesial TLE, 6 lateral TLE) and 2 control groups ( $n = 29$  for FDG,  $n = 20$  for FMZ). EZ of each patient was confirmed by postsurgical pathology and 1-year follow-up, while propagation zone (PZ) and noninvolved zone (NIZ) were derived from the epileptogenicity index based on presurgical stereo-encephalography (SEEG) monitoring. Whole brain PET scans were obtained with dual tracers [ $^{18}\text{F}$ ]FDG and [ $^{18}\text{F}$ ]FMZ on separate days, from which standard uptake value ratio (SUVR) was calculated by global mean scaling. Low-order statistical parameters of SUVRs and t-maps derived against control groups were extracted. Additionally, fused FDG and FMZ features were created by using arithmetic operations. Spearman correlation was used to investigate the associations between FDG and FMZ, while multiple linear regression analyses were used to explore the interaction effects of imaging features in predicting epileptogenicity. Crafted imaging features were used to train logistic regression models to predict EZ, whose performance was evaluated by using 10-fold cross-validation at ROI level, and leave-1-patient-out cross-validation at patient level.

**RESULTS:** FDG SUVR significantly decreased in EZ and PZ compared with NIZ, while FMZ SUVR in EZ significantly differed from PZ. Interaction effects were found between FDG and FMZ in their prediction of epileptogenicity. Fusion of FDG and FMZ provided the best prediction model with an area under the curve (AUC) of 0.86 [0.84–0.87] for EZ versus NIZ and an AUC of 0.79 [0.77–0.81] for EZ versus PZ, eliminating 100% false-positives in 50% of patients, and  $\geq 80\%$  FPs in 90% of patients at patient level.

**CONCLUSIONS:** Combined FDG and FMZ offer a promising avenue for noninvasive localization of the epileptogenic zone in TLE, potentially refining surgical planning.

**ABBREVIATIONS:** AUC = area under the curve; EI = epileptogenicity index; EZ = epileptogenic zone; FCD = focal cortical dysplasia; FMZ = flumazenil; FP = false-positive; GABA<sub>A</sub> = gamma-aminobutyric acid type A; HPC = Human Connectome Project; LASSO = least absolute shrinkage and selection operator; MAD = mean absolute deviation; MNI = Montreal Neurological Institute; NIZ = not-involved zone; PZ = propagation zone; RMS = root mean square; ROC = receiver operating characteristic curve; SEEG = stereo-electroencephalography; SPM = statistical parametric mapping; SUVR = standard uptake value ratio; TLE = temporal lobe epilepsy; TP = true-positive

Epilepsy, affecting around 50 million people worldwide, is characterized by recurrent spontaneous seizures.<sup>1</sup> Approximately 30% of cases are drug-resistant, often requiring surgical resection.<sup>2</sup> Accurate localization of the epileptogenic zone (EZ) is

crucial. The epileptogenicity index (EI), derived from stereo-electroencephalography (SEEG) signals, can be used to classify brain regions as EZ, propagation zone (PZ), or noninvolved zone (NIZ).<sup>3</sup> However, SEEG is costly, samples the brain sparsely, and poses surgical risks.<sup>4</sup>

[ $^{18}\text{F}$ ]FDG-PET noninvasively detects metabolic changes and is frequently utilized when MRI examinations are inconclusive.<sup>5,6</sup>

Received August 6, 2024; accepted after revision January 7, 2025.

From the National Engineering Research Center of Advanced Magnetic Resonance Technologies for Diagnosis and Therapy, School of Biomedical Engineering, (B.C., H.H., J.L., S.Y., Y.C., J.L.), Shanghai Jiao Tong University, Shanghai, China; and Department of Neurosurgery (S.J., J.H., L.C.), and PET Center (W.B.), Huashan Hospital, Fudan University, Shanghai, China.

Bingyang Cai and Shize Jiang are co-first authors. Liang Chen and Jie Luo share senior authorships.

This work was supported by the National Natural Science Foundation of China (No.82272116, No.62101321), and the Science and Technology Commission of Shanghai municipality frontier innovation program (No.24DP3200600).

Please address correspondence to Jie Luo, Room 325, Med-X Institute, 1954, Huashan Rd, Shanghai, China 200030; e-mail: jieluo@sjtu.edu.cn

 Indicates article with supplemental data.

 This article contains an [imaging protocol](#) available on our website.  
<http://dx.doi.org/10.3174/ajnr.A8647>

## SUMMARY

**PREVIOUS LITERATURE:** Invasive SEEG is currently used to capture epileptogenic networks and to identify EZ for patients with drug-resistant epilepsy. [ $^{18}\text{F}$ ]FDG-PET is widely used to identify hypometabolic regions in the epileptic brain, which has been shown to correlate with SEEG-defined EZ. [ $^{18}\text{F}$ ]FMZ-PET targets benzodiazepine sites on GABA<sub>A</sub> receptors, providing insights into inhibitory neurotransmission changes in epilepsy, yet no report had associated FMZ uptake with SEEG-defined epileptogenic networks. A meta-analysis of 34 [ $^{18}\text{F}$ ]FDG studies, 3 [ $^{11}\text{C}$ ]FMZ studies, and 7 combined studies reported comparable lesion detection performances. However, whether image fusion of FDG and FMZ enhances EZ prediction remains unexplored.

**KEY FINDINGS:** [ $^{18}\text{F}$ ]FDG and [ $^{18}\text{F}$ ]FMZ uptake exhibit significant decreases in EZ, while showing an interaction effect in predicting epileptogenicity. Their image fusion features effectively distinguish EZ from propagation and noninvolved zones, achieving an 80% true-positive rate for EZ prediction and reducing false-positives by  $\geq 80\%$  in 90% of patients.

**KNOWLEDGE ADVANCEMENT:** This study pioneers in characterizing FMZ uptake in SEEG-defined epileptogenic networks, uncovering the interaction effect between FDG and FMZ in the epileptic brain. Fusion of FDG and FMZ shows potential to reduce false-positives and enhance accuracy in noninvasive prediction of EZ, offering to enhance diagnostic precision and guide interventions.

Reduced glucose uptake may be attributable to neuronal damage, cumulative mitochondrial stress, or enhanced glycogen storage caused by recurrent seizures.<sup>7</sup> Interictal hypometabolism enhances detection sensitivity in MRI-negative temporal lobe epilepsy (TLE)<sup>8,9</sup> and focal cortical dysplasia (FCD) cases.<sup>10</sup> FDG uptake has been reported to decrease in both SEEG-defined EZ and PZ in patients with FCD.<sup>11</sup>

Flumazenil (FMZ)-based PET, such as [ $^{11}\text{C}$ ]FMZ and [ $^{18}\text{F}$ ]FMZ, target  $\gamma$ -aminobutyric acid receptor A (GABA<sub>A</sub>) receptors, have also been reported to aid the detection of seizure onset zone for drug-resistant TLE, even in MR-negative cases.<sup>6,12,13</sup> GABA<sub>A</sub> receptor-mediated inhibition is a key pathophysiologic mechanism driving increased neuronal excitability and leading to epileptogenesis.<sup>14–17</sup> To date, there is no report on FMZ uptake changes in SEEG-defined epileptogenic networks.

A meta-analysis pooling 34 [ $^{18}\text{F}$ ]FDG studies, 3 [ $^{11}\text{C}$ ]FMZ studies, and 7 [ $^{11}\text{C}$ ]FMZ/[ $^{18}\text{F}$ ]FDG studies reported comparable lesion detection performances of FDG-PET and FMZ-PET.<sup>18</sup> Previous studies have demonstrated that reading both FMZ-PET and FDG-PET scans can assist in EZ localization during clinical presurgical evaluations.<sup>13,19,20</sup> This study hypothesizes that image feature fusion of FDG and FMZ could enhance epileptogenic zone localization accuracy, given their separate pathophysiological mechanisms. We assessed the performance of FDG and FMZ, both separately and combined, in classifying EZ in patients with TLE by using machine learning. We also explored interactions between FDG and FMZ changes to understand their synergy in localizing the EZ. Methods and results adhere to the STARD guidelines (Supplemental Data).

## MATERIALS AND METHODS

### Patient Recruitment

This retrospective study was approved by the Institutional Review Board of Huashan Hospital (IRB No. KY2015-256), and informed consent was obtained from all participants who were patients with drug-resistant TLE being considered for presurgical evaluations between 2018 and 2022. Participants underwent both FDG and FMZ PET scans for clinical evaluations. Within this cohort, were 20 subjects (men/women 11/9, aged 8–52) with SEEG recordings of 2–3 seizures and their EZ location confirmed by postsurgical outcomes and follow-up (Fig 1A).

Three control groups were included for comparison: healthy volunteers for [ $^{18}\text{F}$ ]FDG-PET scans (IRB No. KY2021-454), structural MRI data from the Human Connectome Project's (HCP) 1200 subjects data release (<https://www.humanconnectome.org/study/hcp-young-adult/document/1200-subjects-data-release>), and age-matched patients with drug-resistant epilepsy who were scanned for presurgical evaluation with negative MRI and negative FMZ-PET were selected as the FMZ control group. This choice was primarily constrained by ethical considerations, preventing the recruitment of healthy volunteers for FMZ-PET.

### SEEG Recordings

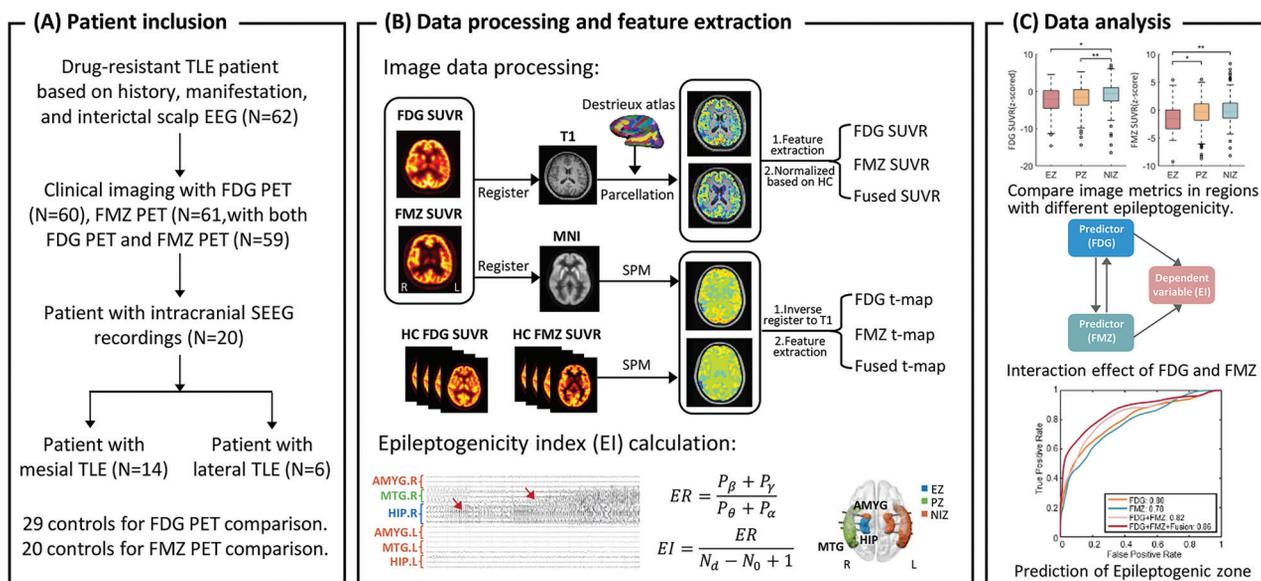
Video-SEEG monitoring employed intracerebral multi-contact electrodes, with 8–16 contacts (2 mm length, 0.8 mm diameter, 1.5 mm spacing). The iEEGview toolbox<sup>21</sup> located each contact anatomically, assigning brain ROIs per the Destrieux atlas.<sup>22</sup> Electrodes in white matter were excluded because of signal interpretation challenges.<sup>23</sup> Bipolar re-referencing minimized common reference and volume conduction effects.<sup>24</sup>

The epileptogenic zone was defined by calculating the EI at each contact based on high-frequency energy ratios relative to seizure onset time.<sup>3</sup> The channel with the highest EI was selected for regions with multiple channels. Brain areas with EI  $> 0.3$  were classified as EZ; those with EI  $\leq 0.3$  and sustained seizure discharge as PZ; all others as NIZ.<sup>3</sup>

### Image Acquisition

All PET imaging was performed on a Biograph mCT Flow Edge 128 scanner (Siemens). [ $^{18}\text{F}$ ]FDG-PET scans were conducted 50 minutes after injecting  $\sim 296\text{MBq}$  (8 mCi) of [ $^{18}\text{F}$ ]FDG lasting 10 minutes. Images were reconstructed with time-of-flight and TrueX algorithm (4 iterations, 21 subsets), smoothed with a 3.5 mm Gaussian kernel, and attenuation-corrected with hybrid CT images. The reconstruction matrix was  $256 \times 256 \times 148$  with a resolution of  $2 \text{ mm} \times 2 \text{ mm} \times 1.5 \text{ mm}$ .

In a separate session, [ $^{18}\text{F}$ ]FMZ-PET scans were acquired (Supplemental Data). [ $^{18}\text{F}$ ]FMZ was synthesized via standard nucleophilic radiofluorination of the corresponding nitro-analog precursor with  $\text{K}^{18}\text{F}$ /kryptofix complex in dimethylformamide (DMF) at



**FIG 1.** Workflow of data collection and processing. A, Flow chart of patient inclusion. B, Workflow of data processing and feature extraction. The SUVR maps of FDG and FMZ images are registered to T1-weighted space and parcellate into 162 ROIs. Individual t-maps of each patient compared with healthy subjects are calculated by using Statistical Parametric Mapping in MNI space and then transformed into individual T1 space. Low-order statistical parameters and fused molecular features are extracted for each ROI from both SUVR maps and t-maps for FDG and FMZ. The epileptogenicity index of the SEEG signals is calculated, with regions subsequently labeled as EZ, PZ, and NIZ based on their epileptogenicity. C, The Mann-Whitney *U* test and the interaction effect analysis are employed to FDG and FMZ SUVR levels across brain regions with different epileptogenicity. Finally, the extracted molecular features are used to build logistic regression for EZ prediction. AMYG, amygdala; HIP, hippocampus; L = left; MTG = middle temporal gyrus; R = right.

160°C for 30 minutes,<sup>25</sup> then purified by high-performance liquid chromatography and sterilized. Scanning began 20 minutes after injecting ~370 MBq (10 mCi) of [<sup>18</sup>F]FMZ, lasting 20 minutes, with reconstruction settings consistent with the FDG session.

MRI structural images were obtained by using a 3D gradient-echo BRAIn Volume (BRAVO; GE Healthcare) sequence with a resolution = 0.9 × 0.9 × 1.0 mm<sup>3</sup>, TR/TE/TI = 8.2/3.2/450 ms, FOV = 240 × 240 × 200 mm<sup>3</sup>, and a 2D T2-weighted FLAIR sequence in 3 orthogonal directions with a resolution = 0.9 × 0.9 mm<sup>2</sup>, slice thickness = 3 mm, TR/TE/TI = 8490/90/2438 ms, FOV = 210 × 300 mm<sup>2</sup>, 46 slices (Supplemental Data).

### Data Processing and Feature Extraction

MRI structural images were parcellated by using Freesurfer with Destrieux atlas, yielding 148 cortical and 14 subcortical ROIs.<sup>22</sup> ROI volumes were calculated after intracranial volume correction and normalized to z-scores by using ComBat-adjusted HCP healthy controls.<sup>26</sup> Standard uptake value ratio (SUVR) maps of PET images were obtained via global mean scaling. Individual t-maps were calculated using statistical parametric mapping (SPM) in Montreal Neurological Institute (MNI) space then transformed into individual T1 space (Fig 1B).

Low-order statistical parameters<sup>27</sup> (mean, median, maximum, minimum, range, standard deviation, variance, root mean square [RMS], mean absolute deviation [MAD], uniformity, skewness, energy, entropy, kurtosis, totaling 14 features) were extracted for each ROI from SUVR maps and t-maps for both FDG and FMZ. SUVR features were further normalized as z-scores based on control groups.

### Feature Fusion

Fused imaging features were engineered through arithmetic operations between FDG and FMZ SUVR features (Addition-s,

Subtraction-s, Product-s, and Logarithm-s, which is the product of logarithms) and their corresponding t-map features (Addition-t, Subtraction-t, Product-t, and Logarithm-t). Additionally, to account for associations between FDG and FMZ, linear fitting parameters between their t-maps, Slope-t and Intercept-t, were included. This resulted in a total of 10 fusion features.

In this study, “fused” specifically refers to fusion features, while “concatenated” describes putting different feature vectors side-by-side to form a new feature vector.

### Statistical Analysis

Sex was compared using the chi-square test, while age was compared using the Mann-Whitney *U* test. To confirm molecular changes in hippocampal sclerosis, the Wilcoxon rank-sum test was used to compare FDG and FMZ SUVR between the ipsilateral and contralateral hippocampi. To evaluate molecular differences among EZ, PZ, and NIZ, we performed analyses by using a linear mixed-effects model to capture group-level differences while controlling for random effects across individuals. Group differences between EZ-PZ, EZ-NIZ, and PZ-NIZ were further examined by using Mann-Whitney *U* tests with effect size calculations. Bonferroni-Holm corrections were applied for multiple comparisons. Additionally, to understand whether FDG and FMZ are correlated in the epileptogenic zone, their correlations in regions of varying epileptogenicity were analyzed by using Spearman correlation. Statistical significance was set at *P* < .05.

### Interaction Analysis

Interaction effects between FDG and FMZ in predicting tissue epileptogenicity were examined through a multiple linear regression model.<sup>28</sup> Predictors included FDG, FMZ, and their interaction

term (FDG × FMZ). Both variables were mean-centered before analysis.<sup>29</sup> A bias-corrected bootstrap approach (5000 iterations) in SPSS Statistics (IBM) tested the significance of interaction effects.

### Epileptogenic Zone Classification

To evaluate whether combining FDG and FMZ imaging could enhance EZ localization, we employed machine learning methods, specifically the least absolute shrinkage and selection operator (LASSO) and logistic regression. The LASSO algorithm was first implemented for feature selection by adding a penalty to the least-squares function to reduce redundancy<sup>30</sup> and enhance predictive power for EI by using 10-fold cross-validation. Logistic regression models then evaluated the capabilities of different modalities in binary classifications of EZ versus PZ, EZ versus NIZ, and PZ versus NIZ with the selected features at both ROI and patient levels. ROI-level classifications used a 10-fold cross-validation and 1000 random trials to minimize random chance. We evaluated the classification efficacy of different logistic regression models that incorporated single-technique features, concatenated features, and concatenated features together with fused imaging features as input. Performance metrics included accuracy, sensitivity, specificity, and the area under the curve (AUC) of receiver operating characteristic (ROC) curves. The Youden index determined optimal cutoff points, and DeLong test compared AUCs across models. Performance was further observed for mesial and lateral TLE groups by using the same model.

At the patient level, the EZ was classified by using leave-1-patient-out cross-validation. A prediction was considered a true-positive (TP) if the model detected a brain region labeled as EZ that was included in the surgical resection area. The false-positive (FP) rate was calculated by combining the results of EZ versus PZ and EZ versus NIZ to find the percentage of contacts falsely assigned as EZ per patient. Patient-level TP and FP were enumerated respectively for mesial and lateral TLE subgroups.

## RESULTS

### Patient Demographics

Patient demographics and clinical information are provided in the Table (see also Supplemental Data), including 14 mesial TLE and 6 lateral TLE. All underwent SEEG and 1-year postsurgery follow-up. Among them, 14 had inconclusive MRI, with 11 bilateral MRI abnormalities and 3 MRI invisible. Patients were compared with healthy controls for FDG-PET (sex,  $P = .99$ ; age,  $P = .08$ ) and the FMZ control group for FMZ (sex,  $P = .75$ ; age,  $P = .12$ ).

**Patient demographics<sup>a</sup>**

Parameter	TLE	HCI	HC2	HCP
Number of subjects	20	29	20	1113
Sex (male/female)	11/9	16/13	10/10	507/606
Age at evaluation (year), median (range)	32 (8–52)	37.5 (11–52)	25 (20–40)	28 (22–36)
Age at seizure onset (year), median (range)	16.5 (1–40)			
Mesial/lateral TLE	14/6			
Epilepsy duration (year), median (range)	9.5 (2–32)			
Seizure frequency (per year), median (range)	30 (1–1277)			
Number of patients with SEEG evaluations	20			
Postsurgical outcome Engel Class (I/II–IV)	19/1			

<sup>a</sup>HCI indicates healthy control group for FDG comparison; HC2, control group for FMZ comparison.

Structural comparisons were also made with healthy HCP subjects (sex,  $P = .40$ ; age,  $P = .12$ ).

### FMZ and FDG PET Show Concordant Findings in Temporal Lobe Epilepsy

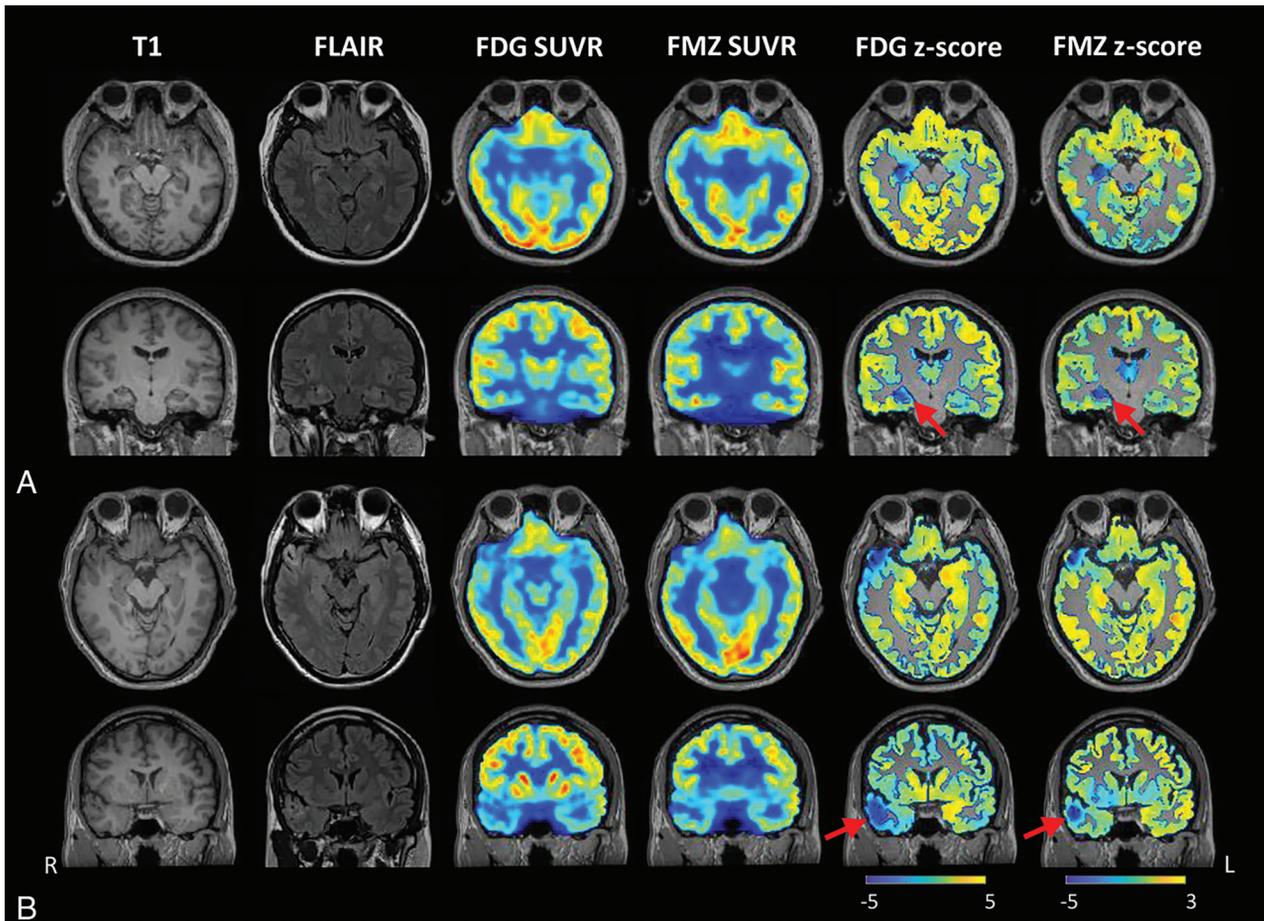
In a representative mesial TLE case, FDG SUVR showed hypometabolism in the right hippocampus, while FMZ SUVR displayed decreased benzodiazepine-GABA<sub>A</sub> receptor binding in the same region (Fig 2A), consistent with FLAIR images. Subgroup comparisons revealed significant differences between the ipsilateral versus contralateral hippocampus in the mesial TLE group ( $P = .003$  for FDG and  $P = .01$  for FMZ) (Supplemental Data). In a lateral TLE, an MRI-positive patient showed reduced FDG and FMZ SUVR in the right inferior temporal gyrus (Fig 2B). However, the lateral TLE group did not exhibit significant differences in the ipsilateral hippocampus ( $P = .16$  for FDG and  $P = .96$  for FMZ) (Supplemental Data).

### Relationships Between FMZ, FDG Uptake, and Epileptogenicity

The epileptogenicity index derived from SEEG classified brain regions into epileptogenic zone, propagation zone, and noninvolved zone. Both FDG and FMZ demonstrated differences among EZ, PZ, and NIZ in the linear mixed-effects model ( $P < .001$  for FDG and  $P < .001$  for FMZ). Both z-scored FDG and FMZ uptakes significantly decrease in EZ compared with NIZ ( $P = .002$ , Cohen  $d = 0.49$ ;  $P < .001$ , Cohen  $d = 0.54$ ). FDG SUVR was decreased in PZ compared with NIZ ( $P < .001$ , Cohen  $d = 0.34$ ), while FMZ SUVR decreased in EZ compared with PZ ( $P = .002$ , Cohen  $d = 0.41$ ) (Fig 3A). The significance threshold was  $\alpha < 0.008$  (0.05/6) after correction for multiple comparisons. Gray matter volume shows a decreasing tendency in EZ compared with NIZ, though not significant after corrections ( $\alpha < 0.017$  [0.05/3], Supplemental Data). FDG and FMZ exhibit stronger positive correlations in EZ than in PZ and NIZ (Fig 3B). These trends were consistent in mesial TLE but not in lateral TLE (Supplemental Data). A significant positive correlation between FDG and FMZ was observed in the EZ of each patient (Fig 3C), indicating an association between glucose uptake and GABA<sub>A</sub> receptor distribution changes within the epileptogenic zone.

### Interaction Effects Between FDG and FMZ

A 2-level multiple linear regression tested the interaction effects between FDG and FMZ in predicting EI. Formula 1 estimates the independent contributions of FDG and FMZ, while formula 2



**FIG 2.** Images of FDG-PET and FMZ-PET in representative cases. A, Axial and coronal views of a patient diagnosed with mesial TLE and pathologically confirmed hippocampal sclerosis (patient #7), with T1-weighted image, FLAIR, [<sup>18</sup>F]FDG-PET, [<sup>18</sup>F]FMZ-PET, and corresponding z-scored maps of FDG and FMZ PET from left to right. B, Axial and coronal views of a patient with lateral TLE in the right inferior temporal gyrus (patient #17).

incorporates an interaction term, “FDG×FMZ.” The analysis showed that FDG and FMZ SUVR independently predicted EI ( $P < .001$ ), with predictive power significantly enhanced by the interaction term ( $\Delta R^2 = 0.03$ ,  $P = .002$ ). When FMZ is used as a conditional value, the negative association between FDG uptake and EI was strong ( $\beta = -0.191$ ,  $P < .001$ ) at low FMZ (1-SD-below the mean) but not significant at medium or high FMZ levels ( $\beta = -0.09$ ,  $P = .08$ , at mean FMZ;  $\beta = 0.02$ ,  $P = .81$ , 1 SD above the mean). Similarly, the negative association between FMZ and EI was strongest at low FDG ( $\beta = -0.35$ ,  $P < .001$ ), moderate at medium FDG ( $\beta = -0.25$ ,  $P < .001$ ), and weakest at high FDG ( $\beta = -0.15$ ,  $P = .03$ ) (Fig 3D).

$$(1) \quad EI = -0.119FDG - 0.295FMZ$$

$$(2) \quad EI = -0.051 - 0.087FDG - 0.250FMZ + 0.103FDG \times FMZ$$

#### Fused FMZ and FDG Provide Improved Prediction of Epileptogenic Zone

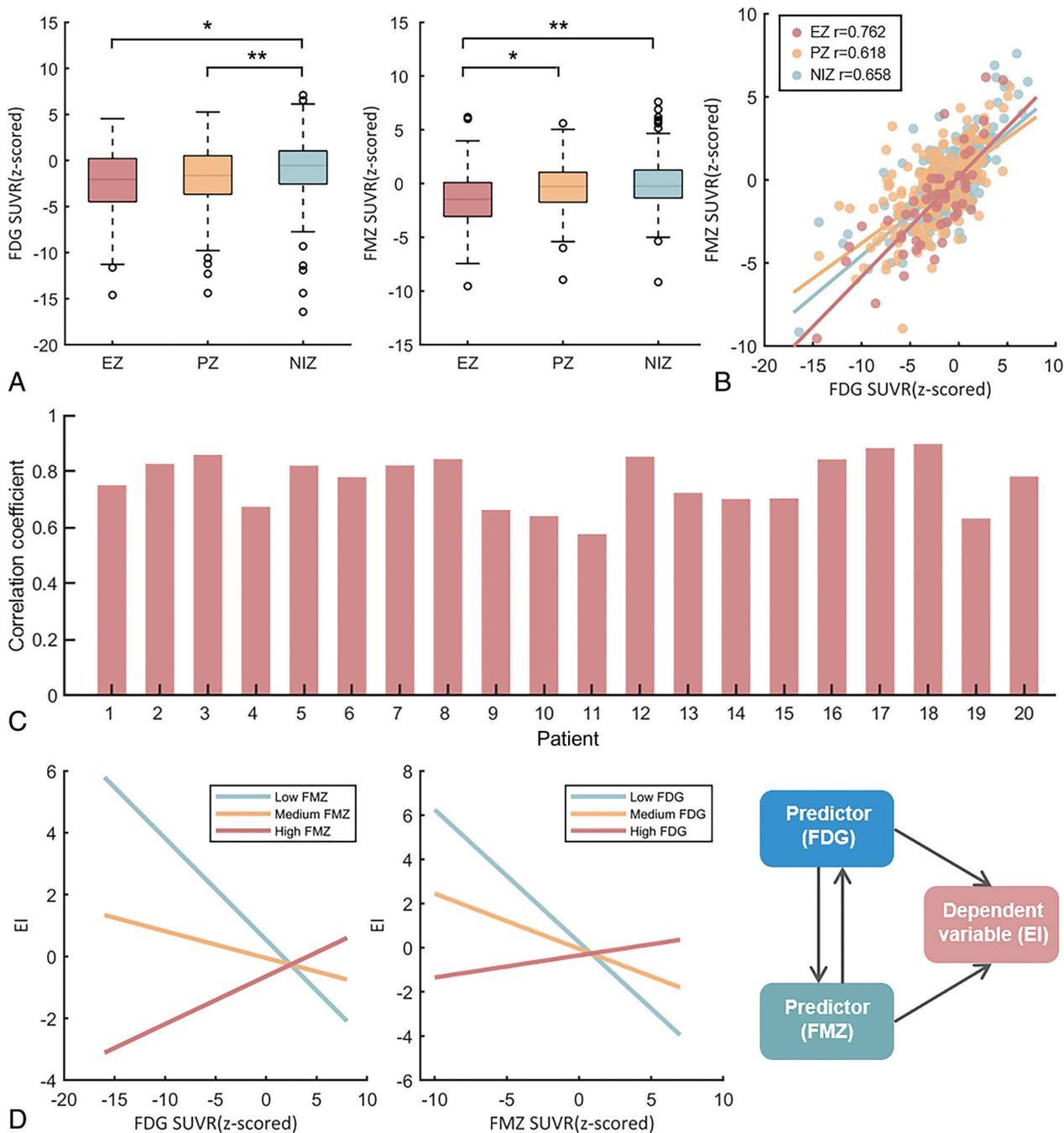
Features were ranked by their LASSO regression coefficient (Fig 4A), with the fusion feature “addition of FMZ and FDG t-maps” showing the highest contribution. For EZ versus NIZ

classifications, FMZ performed comparably with FDG (AUC = 0.78 [0.76–0.79] versus AUC = 0.80 [0.78–0.81]). Their concatenation slightly improved performance (AUC = 0.82 [0.81–0.84]), with fusion features achieving the best results (AUC = 0.86 [0.84–0.87]) (Fig 4B, -C). DeLong test indicates that fusion features significantly outperformed single-technique inputs ( $P = .04$  for FDG,  $P = .008$  for FMZ). EZ versus PZ classifications were more challenging, with PZ versus NIZ separation proving the most difficult (AUC < 0.60). Models predicted EZ more accurately in mesial TLE than in lateral TLE (Supplemental Data).

In patient-level EZ prediction, the concatenated FDG, FMZ, and fusion model showed a similar detection rate as FDG alone in 80% of patients (16/20) (Fig 5A) while eliminating all FPs in 50% of patients and  $\geq 80\%$  of FPs in 90% (18/20) of patients (Fig 5B). Both mesial and lateral TLE subtypes showed similar FP performance, but mesial TLE had a higher true-positive rate compared with lateral TLE (Fig 5A, -C).

#### DISCUSSION

This study demonstrates that [<sup>18</sup>F]FMZ-PET complements the more commonly used [<sup>18</sup>F]FDG-PET in the localization of the epileptogenic zone, reducing false-positives. Significant decreases in both FDG and FMZ uptakes were observed in pathologically confirmed lesions. Multiregression analysis reveals that FDG



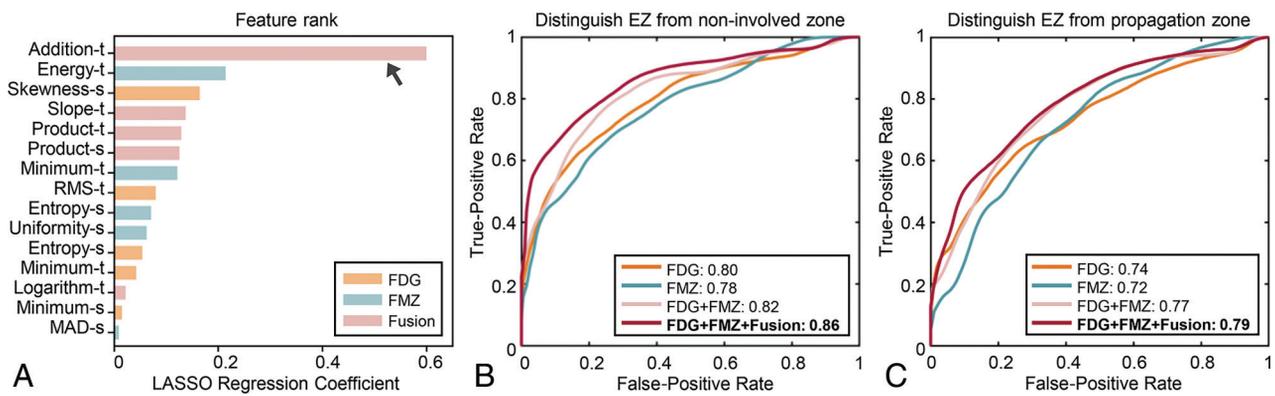
**FIG 3.** The relationship of FDG and FMZ SUVR among brain regions with different epileptogenicity. *A*, Boxplot of z-scored FDG SUVR (*left*) and boxplot of z-scored FMZ SUVR (*right*) in regions with different epileptogenicity noted as EZ, PZ, and NIZ. *B*, The scatterplot of FMZ versus FDG shows their associations change with epileptogenicity. *C*, Correlation coefficients of FMZ and FDG in EZ of all patients. *D*, Interaction effects between FDG and FMZ in predicting epileptogenicity. The relationship between FDG SUVR and EI at 3 FMZ levels (*left*). The relationship between FMZ SUVR and EI at 3 FDG levels (*middle*). Schematic of the interaction model between FDG and FMZ for EI prediction (*right*). \* $P < .05$ , \*\* $P < .01$ , Mann-Whitney  $U$  test under Bonferroni-Holm correction.

SUVR interacts with FMZ SUVR in predicting epileptogenicity, with their fused feature contributing the most in classification models.

### Interplays Between Glucose Uptake and GABA<sub>A</sub> Receptor Availability

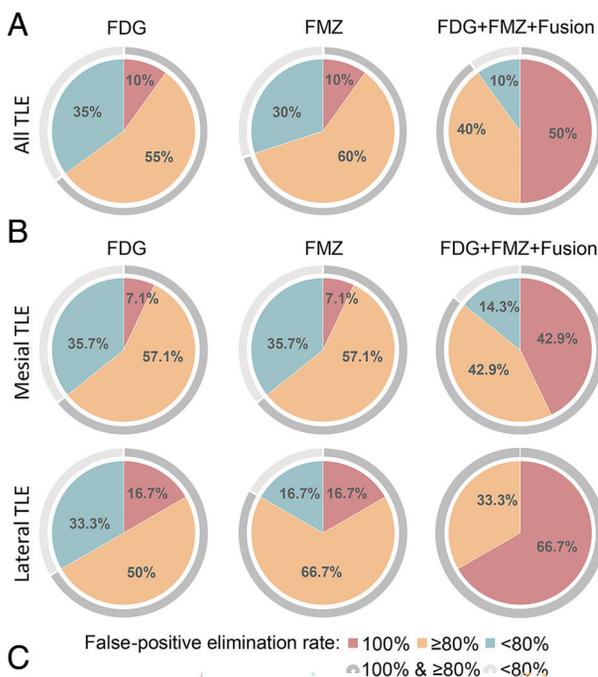
FDG and FMZ SUVR exhibited strong positive correlations in the EZ of each patient and weaker correlations in regions with

lower epileptogenicity. Previous studies comparing FMZ-PET and FDG-PET primarily focused on lesion detection<sup>12</sup> rather than their associations. In temporal lobe epilepsy, reduced glucose metabolism may impair GABA<sub>A</sub> receptor phosphorylation, affecting GABAergic inhibition.<sup>31</sup> In Huntington disease, glucose hypometabolism has been shown to precede GABAergic dysfunction.<sup>32</sup> Conversely, dysfunctional GABA<sub>A</sub> receptors may elevate neuronal activity, altering glucose metabolism by increasing energy



**FIG 4.** Classification of the EZ by using single or concatenated molecular features at ROI level. A, Bar plot showing feature ranking based on LASSO regression coefficients for selected features. The arrow highlights the feature with the highest contribution, “the addition of FMZ and FDG t-maps,” in the LASSO regression model. B, ROC curves of classification models that distinguish EZ from NIZ. C, ROC curves of classification models that distinguish EZ from PZ with 4 different feature inputs. The suffix “-t” means t-map features, while “-s” means SUVR features.

Feature	Number of patients with true positive(%)		
	All subtype (N=20)	Mesial TLE (N=14)	Lateral TLE (N=6)
FDG	16(80)	13(93)	3(50)
FMZ	14(70)	12(86)	2(33)
FDG+FMZ+Fusion	<b>16(80)</b>	<b>13(93)</b>	<b>3(50)</b>



**FIG 5.** Model performance by using single or concatenated molecular features at the patient level. A, The number of patients with true-positive EZ prediction across all subtypes by using FDG (a PET tracer for hypometabolism), FMZ (a PET tracer for GABA<sub>A</sub> receptor binding), and their fusion feature as inputs. B, The patient-level false-positive rate in all subtypes. C, The patient-level false-positive rate in mesial TLE (above) and lateral TLE (below). Different colors represent the range of false-positive elimination: red indicates 100% elimination; orange represents more than 80% elimination; cyan denotes less than 80% elimination; dark gray indicates the combination of 100% and more than 80% elimination; light gray represents less than 80% elimination.

demands.<sup>33</sup> GABAergic neuronal loss would result in decreased glucose uptake and GABA<sub>A</sub> receptor availability.<sup>34</sup> Our analysis reveals for the first time that FMZ SUVR interacts with FDG SUVR in predicting epileptogenicity, indicating their fusion may enhance EZ prediction.

#### Feature Contributions to EZ Classification Models

Machine learning evaluations support synergistic effects between FDG and FMZ. The fused FDG and FMZ features ranked highest in feature importance, which is consistent with the interaction effect analysis. In patient-level classification, the dramatic elimination of false-positives by using fused molecular feature input may be explained by a lack of concurrent hypometabolism and GABA<sub>A</sub> down-regulation in regions with transient functional changes.

Model comparisons revealed better performance for mesial than lateral TLE groups at both ROI level and patient level (Supplemental Data, Fig 5A), which may be attributed to the heterogeneous expression of GABA<sub>A</sub> receptor subunits and variations in lesion locations.

#### Limitations

First, the EZ defined by the epileptogenicity index differs conceptually from the seizure onset zone confirmed by postsurgical seizure freedom. Therefore, for patient-level true-positive detection, we used only the EZ that corroborated the seizure onset zone for each patient. The advantage of using EI is that it provides additional insight into the epileptogenic network, including propagation and noninvolved zones.

Second, due to challenges in recruiting healthy volunteers for FMZ-PET and inherent bias between our data and the [<sup>11</sup>C]FMZ brain template ( $n = 16$ )<sup>35</sup> (eg, demographic and tracer differences), we formed an FMZ control group comprised of radiologically “normal” patients with epilepsy.

Additionally, the small sample size may introduce random variability and bias in machine learning validation, affecting model performance and leading to overfitting. It may also result in an unreliable representation of subgroups, such as lateral versus mesial TLE, potentially skewing results.

Finally, while this study offers valuable insights into the synergistic effects of FDG and FMZ in predicting epileptogenicity, future research should incorporate additional imaging modalities, such as high-resolution structural and functional imaging, to provide a more comprehensive understanding of epileptogenic network alterations in epilepsy.

## CONCLUSIONS

The fusion of FDG and FMZ PET, in conjunction with machine learning techniques, represents a novel and powerful tool for detecting and characterizing the epileptogenic zone in patients with TLE. This approach has potential implications for improving surgical planning and predicting surgical outcomes, thereby contributing to optimizing patient management in epilepsy care.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Beghi E. **The epidemiology of epilepsy.** *Neuroepidemiology* 2020; 54:185–91 CrossRef Medline
2. Zijlmans M, Zweiphenning W, van Klink N. **Changing concepts in presurgical assessment for epilepsy surgery.** *Nat Rev Neurol* 2019;15:594–606 CrossRef Medline
3. Bartolomei F, Chauvel P, Wendling F. **Epileptogenicity of brain structures in human temporal lobe epilepsy: a quantified study from intracerebral EEG.** *Brain* 2008;131:1818–30 CrossRef Medline
4. Mullin JP, Shriver M, Alomar S, et al. **Is SEEG safe? A systematic review and meta-analysis of stereo-electroencephalography-related complications.** *Epilepsia* 2016;57:386–401 CrossRef Medline
5. Lamusuo S, Jutila L, Ylinen A, et al. **[18F]FDG-PET reveals temporal hypometabolism in patients with temporal lobe epilepsy even when quantitative MRI and histopathological analysis show only mild hippocampal damage.** *Arch Neurol* 2001;58:933–39 CrossRef Medline
6. Traub-Weidinger T, Arbizu J, Barthel H, et al. **EANM practice guidelines for an appropriate use of PET and SPECT for patients with epilepsy.** *Eur J Nucl Med Mol Imaging* 2024;51:1891–908 CrossRef Medline
7. Chassoux F, Artiges E, Semah F, et al. **Determinants of brain metabolism changes in mesial temporal lobe epilepsy.** *Epilepsia* 2016;57: 907–19 CrossRef Medline
8. Zhang M, Huang H, Liu W, et al. **Combined quantitative T2 mapping and [(18)F]FDG PET could improve lateralization of mesial temporal lobe epilepsy.** *Eur Radiology* 2022;32:6108–17 CrossRef Medline
9. Huang H, Zhang M, Zhao Y, et al. **Simultaneous high-resolution whole-brain MR spectroscopy and [18F]FDG PET for temporal lobe epilepsy.** *Eur J Nucl Med Mol Imaging* 2024;51:721–33 CrossRef Medline
10. Chassoux F, Rodrigo S, Semah F, et al. **FDG-PET improves surgical outcome in negative MRI Taylor-type focal cortical dysplasias.** *Neurology* 2010;75:2168–75 CrossRef Medline
11. Lagarde S, Boucekine M, McGonigal A, et al. **Relationship between PET metabolism and SEEG epileptogenicity in focal lesional epilepsy.** *Eur J Nucl Med Mol Imaging* 2020;47:3130–42 CrossRef Medline
12. Ryvlin P, Bouvard S, Le Bars D, et al. **Clinical utility of flumazenil-PET versus [18F]fluorodeoxyglucose-PET and MRI in refractory partial epilepsy. A prospective study in 100 patients.** *Brain* 1998;121:2067–81 CrossRef Medline
13. Vivash L, Gregoire MC, Lau EW, et al. **18F-flumazenil: a  $\gamma$ -amino-butyric acid A-specific PET radiotracer for the localization of drug-resistant temporal lobe epilepsy.** *J Nucl Med* 2013;54:1270–77 CrossRef Medline
14. Jansen LA, Peugh LD, Roden WH, et al. **Impaired maturation of cortical GABA<sub>A</sub> receptor expression in pediatric epilepsy.** *Epilepsia* 2010;51:1456–67 CrossRef Medline
15. Deeb TZ, Maguire J, Moss SJ. **Possible alterations in GABA<sub>A</sub> receptor signaling that underlie benzodiazepine-resistant seizures.** *Epilepsia* 2012;53 Suppl 9:79–88 CrossRef Medline
16. Alhourani A, Fish KN, Wozny TA, et al. **GABA bouton subpopulations in the human dentate gyrus are differentially altered in mesial temporal lobe epilepsy.** *J Neurophysiol* 2020;123:392–406 CrossRef Medline
17. Benarroch EE. **GABA<sub>A</sub> receptor heterogeneity, function, and implications for epilepsy.** *Neurology* 2007;68:612–14 CrossRef Medline
18. Niu N, Xing H, Wu M, et al. **Performance of PET imaging for the localization of epileptogenic zone in patients with epilepsy: a meta-analysis.** *Eur Radiology* 2021;31:6353–66 CrossRef Medline
19. Hodolic M, Topkian R, Pichler R. **(18)F-fluorodeoxyglucose and (18)F-flumazenil positron emission tomography in patients with refractory epilepsy.** *Radiology Oncol* 2016;50:247–53 CrossRef Medline
20. Avendaño-Estrada A, Velasco F, Velasco Ana L, et al. **Quantitative analysis of [18F]FFMZ and [18F]FDG PET studies in the localization of seizure onset zone in drug-resistant temporal lobe epilepsy.** *Stereotact Funct Neurosurg* 2019;97:232–40 CrossRef Medline
21. Li G, Jiang S, Chen C, et al. **iEEGview: an open-source multifunction GUI-based Matlab toolbox for localization and visualization of human intracranial electrodes.** *J Neural Eng* 2019;17:016016 CrossRef Medline
22. Destrieux C, Fischl B, Dale A, et al. **Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature.** *Neuroimage* 2010;53:1–15 CrossRef Medline
23. Mercier MR, Dubarry AS, Tadel F, et al. **Advances in human intracranial electroencephalography research, guidelines and good practices.** *Neuroimage* 2022;260:119438 CrossRef Medline
24. Lagarde S, Roehri N, Lambert I, et al. **Interictal stereotactic-EEG functional connectivity in refractory focal epilepsies.** *Brain* 2018; 141:2966–80 CrossRef Medline
25. Ryzhikov NN, Seneca N, Krasikova RN, et al. **Preparation of highly specific radioactivity [18F]flumazenil and its evaluation in cynomolgus monkey by positron emission tomography.** *Nucl Med Biol* 2005;32:109–16 CrossRef Medline
26. Radua J, Vieta E, Shinohara R; ENIGMA Consortium Collaborators, et al. **Increased power by harmonizing structural MRI site differences with the ComBat batch adjustment method in ENIGMA.** *Neuroimage* 2020;218:116956 CrossRef Medline
27. Rizzo S, Botta F, Raimondi S, et al. **Radiomics: the facts and the challenges of image analysis.** *Eur Radiology Exp* 2018;2:36 CrossRef Medline
28. Garofalo S, Giovagnoli S, Orsoni M, et al. **Interaction effect: are you doing the right thing?** *PLoS One* 2022;17:e0271668. CrossRef Medline
29. Iacobucci D, Schneider MJ, Popovich DL, et al. **Mean centering helps alleviate “micro” but not “macro” multicollinearity.** *Behav Res Methods* 2016;48:1308–17 CrossRef Medline
30. Tibshirani R. **Regression shrinkage and selection via the LASSO.** *J R Stat Soc Series B Stat Methodol* 1996;58:267–88 CrossRef
31. Laschet JJ, Kurcewicz I, Minier F, et al. **Dysfunction of GABA<sub>A</sub> receptor glycolysis-dependent modulation in human partial epilepsy.** *Proc Natl Acad Sci U S A* 2007;104:3472–77 CrossRef Medline
32. Holthoff VA, Koeppe RA, Frey KA, et al. **Positron emission tomography measures of benzodiazepine receptors in Huntington’s disease.** *Ann Neurol* 1993;34:76–81 CrossRef Medline
33. Nedergaard S, Andreassen M. **Opposing effects of 2-deoxy-d-glucose on interictal- and ictal-like activity when K(+) currents and GABA (A) receptors are blocked in rat hippocampus in vitro.** *J Neurophysiol* 2018;119:1912–23 CrossRef Medline
34. Kim K, Yoon H. **Gamma-aminobutyric acid signaling in damage response, metabolism, and disease.** *Int J Mol Sci* 2023;24:4584 CrossRef Medline
35. Nørgaard M, Beliveau V, Ganz M, et al. **A high-resolution in vivo atlas of the human brain’s benzodiazepine binding site of GABA (A) receptors.** *Neuroimage* 2021;232:117878 CrossRef Medline

# MRI Interpretation Errors in Adult Patients with Medically Refractory Epilepsy

 Aoife M. Haughey,  Nadav Gasner, and  Timo Krings



## ABSTRACT

**BACKGROUND AND PURPOSE:** The aim is to determine the most common MRI interpretation errors in patients with medically refractory epilepsy (MRE) according to our experience in a tertiary academic center with a large volume epilepsy program, to raise awareness of the “blind spots” in imaging patients with MRE, and to highlight the importance of combining clinical and electroencephalographic information to obtain the most accurate diagnosis.

**MATERIALS AND METHODS:** This is a retrospective observational study. All patients with MRE who underwent MRI brain (on 3T, with dedicated epilepsy protocol), and who were discussed at the weekly interdisciplinary epilepsy conference in our center between January 2008 and July 2023 were included. The initial MRI interpretation and final MRI interpretation results were reviewed.

**RESULTS:** A total of 886 patients with MRE were included. Three hundred patients were MRI-negative (33.86%), diagnoses were missed in 95 patients (10.7%), a second diagnosis was missed in patients with dual pathologies in 42 patients (4.74%), findings were misinterpreted in 4 patients (0.45%), an overcall was made in 2 patients (0.2%), and in 5 patients the correct diagnosis was discussed but erroneously deemed absent on initial interpretation (0.56%), resulting in a total of 148 discrepant MRI reports. Mesial temporal sclerosis comprised the most common pathology encountered overall, followed by encephalomalacia related to prior insult, enlarged amygdala (EA), malformations of cortical development, cavernoma, and ulegyria. The relative proportion of missed focal cortical dysplasia, encephalocele, and EA accounted for the most common misses/misdiagnoses.

**CONCLUSIONS:** Evaluation of patients with MRE requires excellent interdisciplinary care. Input from all members of the interdisciplinary team is essential for accurate interpretation of MRI in patients with MRE for the neuroradiologist. We hope to inform radiologists of commonly overlooked pathologies in MRI brain interpretation for patients with epilepsy. In doing so, we want to maximize the yield of initial MRI interpretation in these patients.

**ABBREVIATIONS:** EA = enlarged amygdala; EEG = electroencephalogram; EMR = electronic medical record; FCD = focal cortical dysplasia; MCD = malformations of cortical development; MEG = magnetoencephalography; MRE = medically refractory epilepsy; MTS = mesial temporal sclerosis; PVNH = periventricular nodular heterotopia; SCN H = subcortical nodular heterotopia; TLE = temporal lobe epilepsy

Epilepsy is a chronic seizure disorder with recurrent and unprovoked seizure activity leading to interruptions of normal brain function.<sup>1</sup> Epilepsy is not a homogeneous disease but rather is an umbrella term given to patients who present with a predisposition

to recurrent and unpredictable seizure activity. There are myriad pathologies and disease processes that may cause seizures, and while all patients with epilepsy have seizures, not all patients who experience seizures have epilepsy.<sup>2</sup>

Epilepsy affects 0.5%–1% of the global population<sup>3</sup> with a cumulative annual incidence of approximately 68 per 100,000.<sup>4</sup> Both incidence and prevalence are marginally higher in males versus females.<sup>1,4</sup> A meta-analysis by Fiest et al<sup>4</sup> found higher incidence in youngest and oldest age groups. Focal seizures represent the most common form of seizure in all age groups.<sup>1</sup> The underlying etiology of epilepsy is varied; however, no cause is identified in up to 50% of patients from high-income countries.<sup>1,5</sup>

Up to 75% of patients with epilepsy are effectively managed with antiepileptic medications.<sup>6</sup> Unfortunately, 25% of patients are diagnosed with medically refractory epilepsy (MRE).<sup>7,8</sup> MRE is defined as “failure of adequate trials of 2 tolerated, appropriately

Received September 9, 2024; accepted after revision December 26.

From the Division of Neuroradiology, Joint Department of Medical Imaging, University Health Network and Toronto Western Hospital, Department of Medical Imaging (A.M.H., N.G., T.K.), University of Toronto, Toronto, Ontario, Canada; Division of Neurology (N.G.), Children’s Hospital of Eastern Ontario, Ottawa, ON, Canada; Division of Neurointerventional Radiology (T.K.), Lahey Hospital and Neurovascular Center Beth Israel Lahey Health, University of Massachusetts Chan Medical School, Boston, Massachusetts; and Edward Singleton Department of Radiology (T.K.), Texas Children’s Hospital and Baylor College of Medicine, Houston, Texas.

Please address correspondence to Dr. Timo Krings, Chair of the Division of Neurointerventional Radiology, Lahey Hospital and Neurovascular Center Beth Israel Lahey Health, University of Massachusetts Chan Medical School, Boston, Massachusetts, 41 Mall Road, Burlington, MA 01805; e-mail: timo.krings@lahey.org; @LaheyRadiology

 Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8681>

## SUMMARY

**PREVIOUS LITERATURE:** MRI plays a critical role in the diagnosis and management of MRE. Previous studies highlight the importance of epilepsy-specific imaging protocols and expert neuroradiology interpretation, particularly for subtle lesions like MTS and FCD. Missed diagnoses on initial MRI interpretation remain common, even in high-volume epilepsy centers, with misinterpretations often attributed to lack of clinical or EEG information. The literature underscores that up to 33% of epileptogenic lesions are overlooked on standard MRI, particularly when not reviewed by subspecialized neuroradiologists. Advanced imaging protocols and multidisciplinary input are key to optimizing diagnostic accuracy.

**KEY FINDINGS:** Among 886 patients with MRE, 16.7% of initial MRI reports were discrepant, with MTS, FCD, and enlarged amygdala being the most frequently missed pathologies. Discrepancies occurred equally across junior and senior radiologists, and protocol changes introduced in 2019 did not significantly reduce the proportion of missed diagnoses in this cohort.

**KNOWLEDGE ADVANCEMENT:** This study highlights common diagnostic pitfalls in epilepsy MRI interpretation and underscores the importance of integrating clinical and EEG data with imaging findings to enhance diagnostic accuracy. It advocates for a standardized approach to epilepsy reporting and the adoption of emerging technologies, such as artificial intelligence and post-processing tools, to assist radiologists in detecting subtle pathologies.

chosen and used antiepileptic drug schedules (whether as monotherapies or in combination) to achieve sustained seizure freedom.<sup>8,9</sup>

Understanding the underlying causes of MRE is essential for guiding appropriate management. Several factors can contribute to development of MRE, including genetic predispositions, specific etiologies like malformations of cortical development, and acquired lesions such as mesial temporal sclerosis (MTS) or enlarged amygdala (EA), resulting from prolonged seizures.<sup>5</sup>

In this cohort, specific and optimized imaging protocols are required for optimal diagnosis. Von Oertzen et al<sup>6</sup> illustrated the benefits of epilepsy-tailored MRI, and of MRI interpretation by a specialist neuroradiologist at a tertiary hospital, in terms of maximizing detection of epileptogenic lesions.<sup>10</sup> They determined that approximately 33% of epileptogenic lesions were overlooked on standard MRI, acquired at 1.5T, that did not include epilepsy-dedicated sequences, such as those optimized for assessing the mesial temporal lobes or focal cortical dysplasia (FCD). Detection rate was also reduced in MRIs interpreted by nonspecialized radiologists.<sup>6</sup> Still some subtle epileptogenic foci are missed even when subspecialty trained radiologists interpret optimized 3T epilepsy protocol MRI.

In this article we aim to determine the most common MRI interpretation errors in patients with MRE according to our experience in a tertiary academic center with a large volume epilepsy program. We aim to raise awareness of “blind spots” in imaging patients with MRE and to highlight the importance of combining clinical information and electroencephalogram (EEG) findings to obtain the most accurate diagnosis.

## MATERIALS AND METHODS

### Study Cohort

This study was exempt from research ethics board review as determined by the University Health Network Quality Improvement Review Committee (QIRC QI ID 21–0211).

This is a retrospective observational study. The design and reporting of this study conform to published Strengthening the Reporting of Observational Studies in Epidemiology guidelines. The cohort comprised all patients diagnosed with MRE, who had

an MRI and who were discussed at the weekly interdisciplinary epilepsy conference at Toronto Western Hospital, between January 2008 and July 2023. In this conference, epilepsy neurologists, neuropsychologists, as well as magnetoencephalography (MEG) and EEG specialists, a neurosurgeon subspecialized in epilepsy surgery, and a neuroradiologist with more than 20 years' experience in reading epilepsy MRIs discuss patients, integrate clinical, radiologic, and EEG information (including intraictal EEG obtained from the epilepsy monitoring unit and MEG) to formulate the most likely cause of a patient's seizures and to discuss management strategies (list of diagnoses, Supplemental Data). Patients were excluded from the study if they did not have at least 1 brain MRI study performed at our institution (excluding patients with imaging performed only in an external institution).

This observational study captured all patients in whom discrepant imaging findings were found as compared with the original report. Following consensus within the multidisciplinary case conference where additional data such as MEG, EEG, and seizure semiology were discussed, imaging findings were reviewed with the primary reading radiologist and an addended consensus report was issued, which for the sake of this article was considered the final diagnosis. We did not have histopathology to confirm the consensus read in most cases.

Discrepancies or errors were subdivided into 5 categories: 1) miss (Supplemental Data), for cases in which a pathology was missed on the initial study/initial MRI interpretation; 2) dual or multiple, for cases in which only part of an imaging diagnosis was made in patients with more than 1 pathology on imaging (Supplemental Data); 3) misinterpretation, for patients in whom an erroneous imaging diagnosis was made due to misinterpretation of the imaging findings; 4) overcall, for patients in whom initial MRI interpretation raised the possibility of an imaging diagnosis, which was ultimately discarded on consensus read; and 5) confirmation, a category reserved for patients in whom the initial interpretation discussed an imaging diagnosis without ultimately making the diagnosis, only for this diagnosis to be confirmed on re-evaluation of MRI or on repeat MRI. An example of this would include reporting of “questionable increased

signal intensity without volume loss within the left hippocampus, which may be seen in MTS, but appearances are nonspecific"; only for this patient to have confirmed MTS on subsequent MRI with congruent EEG and clinical findings.

### **MRI**

All patients had at least 1 3T MRI performed at our institution, with most patients scanned according to a dedicated epilepsy protocol on either a Signa HDxt (GE Healthcare) or a Skyra (Siemens Healthineers).

Imaging parameters for our epilepsy protocol were as follows. For GE Healthcare, axial FLAIR: TR/TE 9102.00/141.17 ms, FOV 22.0 × 22.0 cm, matrix 384 × 224, slice thickness 4.0 mm; coronal oblique T2 perpendicular to the hippocampus: TR/TE 6500/40 ms, FOV 22.0 × 22.0 cm, matrix 512 × 512, slice thickness 3.0 mm; coronal oblique FLAIR perpendicular to the hippocampus: TR/TE 8802/143.34 ms, FOV 22.0 × 22.0 cm, matrix 252 × 224, slice thickness 4.0 mm; sagittal 3D T1: TR/TE 7.13/2.94 ms, FOV 22.0 × 22.0, matrix 256 × 256, slice thickness 1.0 mm; DWI: TR/TE 8000.00/81.90 ms, FOV 22.0 × 22.0 cm, matrix 256 × 256, slice thickness 4.0 mm; GRE: TR/TE 2000/25 ms, FOV 22.0 × 22.0 cm, matrix 264 × 256, slice thickness 4.0 mm.

For Siemens Healthineers, axial and coronal oblique FLAIR perpendicular to the hippocampus: TR/TE 9000/94 ms, FOV 19.9 × 22.0 cm, matrix 256 × 232, slice thickness 4.0 mm; coronal oblique T2 perpendicular to the hippocampus: TR/TE 6900/28 ms, FOV 22.0 × 18.6 cm, matrix 432 × 512, slice thickness 3.0 mm; sagittal 3D T1: TR/TE 2300/2.27 ms, FOV 25.0 × 25.0, slice thickness 1.0 mm, matrix 256 × 256; DWI: TR/TE 6900/94 ms, FOV 22.0 × 22.0 cm, matrix 160 × 160, slice thickness 4.0 mm; SWI: TR/TE 28/20 ms, FOV 18.0 × 23.0 cm, matrix 320 × 230, slice thickness 2.6 mm. From 2019 onwards, 3D FLAIR (1 mm) and high-resolution 2D coronal T2-weighted sequences (0.4 × 0.4 × 2 mm) were included in the protocol.<sup>11</sup>

No postprocessing was used in image interpretation.

### **RESULTS**

A total of 886 patients with MRE were identified; 459 female and 427 male patients were included with a mean age of 35.7 years and age range of 18–79 years. In this cohort, 300 of 886 patients were MRI-negative (33.86%), diagnoses were missed (miss) in 95 of 886 patients (10.7%), a second diagnosis was missed in patients with dual or multiple pathologies in 42 of 886 patients (dual) (4.74%), findings were misinterpreted in 4 of 886 patients (misinterpretation) (0.45%), diagnoses were overcalled in 2 of 886 patients (overcall) (0.2%), and in 5 of 886 patients the correct diagnosis was discussed but erroneously reported as absent (confirmation) (0.56%), resulting in a total of 148 of 886 (16.7%) discrepant MRI reports. The remaining 438 of 886 (49.44%) cases were correctly identified on imaging, including patients with correctly identified dual and multiple pathologies.

A total of 20 subspecialist fellowship-trained neuroradiology staff reported the index MRI, with variable years in practice ranging from 1 to more than 40 years' experience. Senior staff (those with 10 years or more of experience working as subspecialist neuroradiologist) reported 770 of 886 (86.9%) of the studies. Junior staff (those with less than 10 years' experience) reported 116 (13.1%) of

the studies. Of the 148 patients with discrepant MRI reports, 126 of 148 were reported by senior staff (85.1%) indicating 16.4% of the total 770 reported by senior staff, while 22 of 148 of the discrepant reads were reported by junior staff (14.9%), indicating 15.5% of the 116 reports made by junior staff.  $\chi^2$  test revealed no statistically significant difference in error rates between the senior and junior radiologists,  $P = .571$ , with statistical significance set at  $P < .05$ .

Of the 148 discrepant reports, 7 (4.7%) were initially read by the subspecialist neuroradiologist with less than 20 years' experience in reading epilepsy, while 210 (23.7%) of the total patient cohort of 886 were reported by this individual. All identified discrepancies were reported by a single subspecialist neuroradiologist expert in epilepsy.

MRI was performed in 635 patients before the implementation of the revised epilepsy MRI protocol that incorporated the recommended 3D FLAIR (1 mm) and high-resolution 2D coronal T2-weighted sequences (0.4 × 0.4 × 2 mm) as recommended by the International League Against Epilepsy,<sup>12</sup> while the remaining 251 had MRI performed after this protocol change. Ninety-five of the discrepancies occurred before revision of the epilepsy protocol in 2019, while 53 occurred thereafter. The 3D FLAIR sequence is optimized for detection of cortical abnormalities. On inspection of our cohort, of the 19 patients in whom FCD was missed, 16 were missed before the introduction of the revised protocol, while 3 were missed thereafter. The high in-plane resolution 2D coronal T2-weighted MRI is optimized for assessment of the hippocampal structures. Reviewing MTS in our cohort, unilateral MTS was diagnosed in 233 patients, of whom 143 patients were diagnosed before the epilepsy MRI protocol change and 90 thereafter. Of the 15 missed diagnoses of MTS, 9 were imaged before and 6 after the introduction of the revised protocol.

The 2-proportion Z-test was used to compare the proportions of missed diagnoses relative to the total number of cases for FCD and MTS before and after the protocol change. This approach ensured that differences in the total number of cases before and after the protocol change were accounted for. The resulting Z-statistic and  $P$  value assess whether the observed difference in proportions is statistically significant. Using the 2-proportion Z-test (Z-statistic:  $-0.041$  and  $P$  value: 0.967 for FCD, and Z-statistic:  $-0.113$  and  $P$  value: 0.910 for MTS), there was no statistically significant difference in the proportion of correct FCD or MTS diagnoses before and after the protocol change.

### **Inadequate Imaging**

For 8 patients in whom the diagnosis was missed, the initial brain MRI was performed at 1.5T. All these patients harbored MTS as subsequently identified with a dedicated subsequent epilepsy protocol at 3T. In 1 further patient in whom a diagnosis of MTS was initially missed on routine brain MRI protocol 3T, unilateral MTS was subsequently detected on a follow-up dedicated epilepsy protocol at 3T.

### **Additional Clinical or EEG Information**

In 39 of 148 patients, due to availability of additional clinical or EEG localizing information made available to the reviewing

neuroradiologist during multidisciplinary rounds, the final diagnoses could be obtained.

In the remaining 109 cases, no new EEG localizing information was obtained between the initial erroneous MRI interpretation and final consensus read. Localizing information was omitted from the index MRI requisition. Consensus diagnosis for these patients was reached by multidisciplinary discussion, incorporating all available information including clinical and EEG findings that in combination contributed to the final diagnosis. EEG findings were deemed notable in our cohort in patients in whom the recorded seizure onset zone on EEG correlated with abnormality seen on the reviewed MRI by our expert radiologist.

### **Pathologies Missed**

The most common overall diagnoses in our cohort included MTS (unilateral,  $n = 233$ , bilateral,  $n = 33$ ), followed by encephalomalacia related to prior insult ( $n = 103$ ), EA ( $n = 63$ ), malformations of cortical development (FCD,  $n = 56$ ; polymicrogyria,  $n = 15$ ; periventricular nodular heterotopia [PVNH],  $n = 19$ ; band heterotopia,  $n = 6$ ; subcortical nodular heterotopia [SCNH],  $n = 2$ ), cavernoma ( $n = 32$ ), and ulegria ( $n = 12$ ). A full breakdown of missed pathologies, of entities missed in those with dual/multiple pathologies and a full list of all interpretation errors is available in the Supplemental Data.

Overall, unilateral MTS was missed in 15 of 266 patients (5.6% of all patients with MTS) as an isolated abnormality. Unilateral MTS accounted for 9 of 42 (21.4%) patients in the dual/multiple error group. Bilateral MTS was missed in 4 of 33 (12%) of all patients with bilateral MTS, misinterpreted as unilateral in 6 of 33 (18.2%) patients with bilateral MTS, and accounted for 6 of 42 (14.2%) patients in the dual/multiple error group.

EA was missed in 26 of 63 patients (41.3%) of all patients with EA and accounted for a further 9 of 42 (14.3%) patients in the dual/multiple category. FCD was missed in 19 of 56 (33.9%) of all patients with FCD and accounted for 8 of 42 (14.3%) in the dual/multiple category on initial read. Encephaloceles were missed in 8 of 26 (30.8%) of all patients with encephaloceles and accounted for 2 of 42 (7.7%) overlooked in the dual/multiple category.

Polymicrogyria was missed in 5 of 17 (29.4%) of all patients with polymicrogyria. PVNH was missed in 2 of 20 (10%) of all affected patients, 1 of 6 (16.7%) band heterotopia, and 1 of 3 (33%) SCNH were missed on initial interpretation. An additional case of band heterotopia accounted for (1/42, 2.4%) overlooked in the dual/multiple category.

Of the 148 patients for whom discrepant MRI reads were detected, 17 demonstrated a generalized seizure pattern on EEG, 125 patients demonstrated focal/localized seizure pattern on EEG, 1 patient was diagnosed with generalized pseudoseizure, and 5 patients had no epileptiform activity detected on EEG.

Of those 42 patients in whom a dual error was made, generalized seizures were detected in 4 patients, while focal seizures were noted in the remaining 38 patients on EEG. In 13 of 42 patients with dual pathology, the second pathology diagnosed on review of MRI colocalized to the same lobe and was concordant with EEG findings, thus deemed clinically significant.

### **DISCUSSION**

In our study of 886 patients with MRE, 300 patients were MRI-negative (33.86%).

Blumcke et al<sup>12</sup> performed a large-scale study in which they assessed the neuropathologic diagnoses of the lesions underlying MRE in nearly 10,000 patients. Similar to our patient cohort, MTS was the most common histopathologic diagnosis, followed by tumors, malformation of cortical development, vascular malformations, glial scars, and encephalitis in order of frequency. Their study found no abnormality in only 8.4% of patients,<sup>12,13</sup> which may be related to selection bias of the investigated cohort.

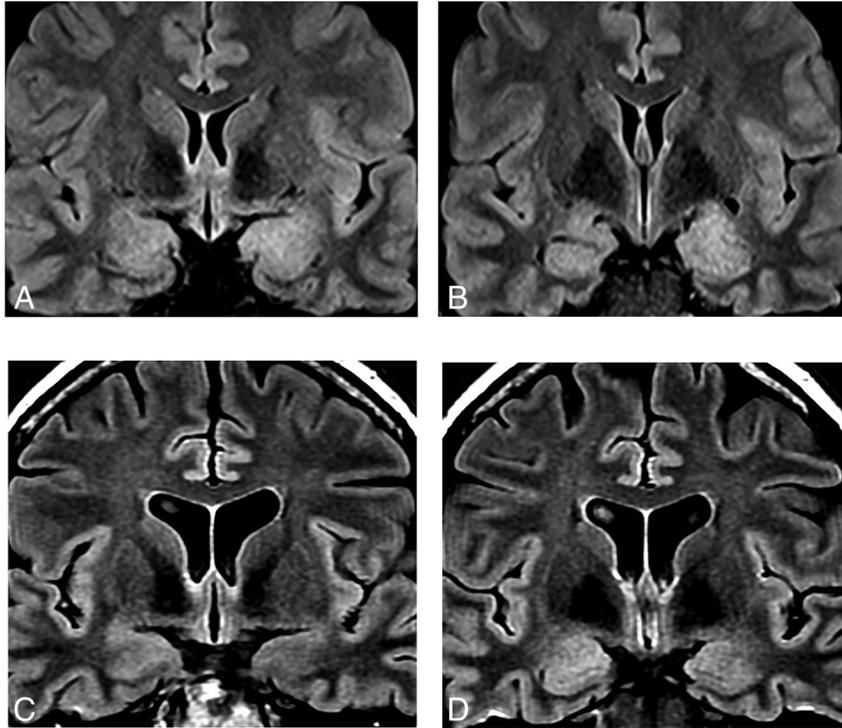
The percentage of MRI-negative patients in our cohort (33.86%) is at the higher end of the previously reported spectrum that ranges from 17%–43%,<sup>10,14,15</sup> which may be related to the fact that in our multidisciplinary epilepsy rounds not only patients with partially complex seizures scheduled for lesionectomy but also patients with generalized seizures investigated for alternate treatment modalities (eg, vagus nerve stimulation) are discussed.

Many series have specifically studied MRI-negative MRE patients.<sup>16,17</sup> Temporal lobe epilepsy (TLE), the most common form of focal epilepsy, has a high chance of progressing to MRE, and as such is often evaluated extensively toward possible surgical intervention.<sup>16</sup> The literature reports that approximately 30% of patients with TLE have normal preoperative MRI<sup>10,16,18</sup> or may have subtle structural lesions overlooked on imaging if not interpreted in conjunction with additional information, such as seizure semiology or EEG.<sup>18-21</sup> Our figure of 33.86% mirrors this figure more closely.

While MTS comprised the most common missed pathology we encountered overall (27/266), its relative proportion was only 10.2% of all cases with MTS, representing a relatively low percentage, while the relative proportion of missed EA (41.3%; 26/63), FCD (33.9%; 19/56), and encephaloceles (14.3%; 8/56) was significantly higher thus constituting the 3 most commonly missed diagnoses in patients with MRE. An additional 15 MTS (5.6%), 8 FCD (14.3%), 2 encephaloceles (7.7%), and 9 EA (14.3%) were missed as the second diagnosis in patients with dual pathology on MRI. Dual pathology was missed in a total of 42 of 148 patients (28.4%).

On review of the breakdown of seniority of those cases in which there were discrepant initial reads, we did not find a statistically significant difference between senior and junior radiologists, underscoring the challenge of detecting subtle lesions, regardless of experience.

Given the retrospective nature of this study, it is difficult to say with certainty the cause for the initial discrepant reads. We consider lack of clinical information a contributing factor. The seizure semiology and/or EEG findings are generally omitted from the MRI requisition. This is often available within subsections of the electronic medical record (EMR). We cannot retrospectively quantify use of EMRs by the initial interpreting radiologist. We understand from local discussion that this information is not routinely sought during MRI interpretation in our institution due to time constraints in daily practice. Temporal changes and increasing demands on institutional and individual workloads are widely recognized. This pressure to read cases quickly is considered a possible factor in erroneous reporting,



**FIG 1.** Coronal FLAIR images from 2 patients (patient 1, A and B, patient 2, C and D, respectively) with evidence of EA. Top images demonstrate FLAIR hyperintensity and enlargement of the left amygdala, while bottom images demonstrate FLAIR signal hyperintensity and enlargement of the right amygdala.

although not quantifiable. Thorough evaluation, especially for small cortical dysplasias or subtle hippocampal sclerosis, demands a considerable investment of time.<sup>11</sup> A formal Relative Value Unit (RVU) model does not apply to our practice.

In the following, we focus on the most overlooked entities to highlight their imaging features to prevent radiologists from overlooking these entities.

### EA

EA (Fig 1) is increasingly recognized as an epileptogenic entity in patients with TLE. The exact etiology of this entity, and its significance in patients with TLE has yet to be fully elucidated.<sup>22</sup>

In the late 1990s, EA was described and studied in patients with mood disorders in conjunction with TLE.<sup>23</sup> Subsequently, it was further investigated in patients with TLE without emphasis on mood imbalances.<sup>24,25</sup>

It has been posited as an epileptogenic focus in patients previously identified as having MRI-negative TLE,<sup>25,26</sup> has been denoted as a seizure onset and irritative zone by others,<sup>27</sup> observed in healthy controls,<sup>26</sup> and has also been suggested to represent seizure-related structural changes in patients with TLE-EA.<sup>28</sup> Some suggest that persistent EA on imaging in patients with epilepsy was correlated with higher incidence of MRE. While other studies suggest it may relate to an underlying autoimmune encephalitis.<sup>29</sup> Resected EA has shown no abnormality on histopathology in some studies,<sup>27</sup> while others documented glioneuronal hamartoma,<sup>30</sup> microdysgenesis, and gliosis<sup>31</sup> as a pathologic correlate of EA. Studies have also assessed for an

association between EA and postoperative outcome,<sup>32</sup> with results suggesting no significant association between EA and postoperative outcome, or between EA on preoperative MRI and histopathologic findings in the resected amygdala.

As may be apparent from this discussion, a consensus as to the exact pathophysiology of EA in TLE has yet to be reached. It is however acknowledged that while EA may represent a seizure-induced structural phenomenon on MRI, it is nonetheless present in patients with TLE, independent of seizure activity. This supports the hypothesis that it may represent an epileptogenic focus, making it an important entity to document on MRI. Because it is a relatively new and poorly understood entity, many radiologists may not be familiar with this imaging diagnosis, thus the high percentage of missed EAs in our cohort.

### Malformations of Cortical Development

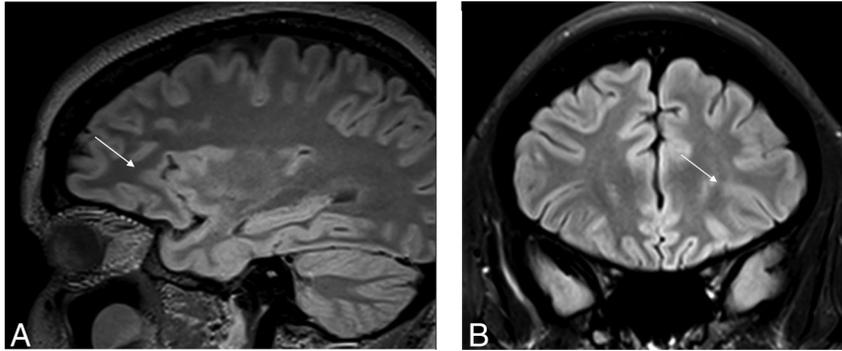
There are many different types of malformations of cortical development (MCD). This umbrella of diagnoses accounts for up to 40% of pediatric epilepsy cases. It is less common in adults but remains an important diagnosis in MRE.

By far the most common MCD encountered on MRI in our cohort was FCD type IIb, documented in 56 patients (6.3%). FCD was missed in 27 of 56 patients (48.2%). Less commonly observed MCD included other abnormalities of neuronal migration, including heterotopia, and abnormalities of post-migrational development including polymicrogyria, schizencephaly, and type I, IIa, and type III FCD. FCD is renowned for being difficult to appreciate on MRI.<sup>33</sup> Reports suggest up to 72% of MRI-negative patients with epilepsy have FCD on histopathology.<sup>33</sup>

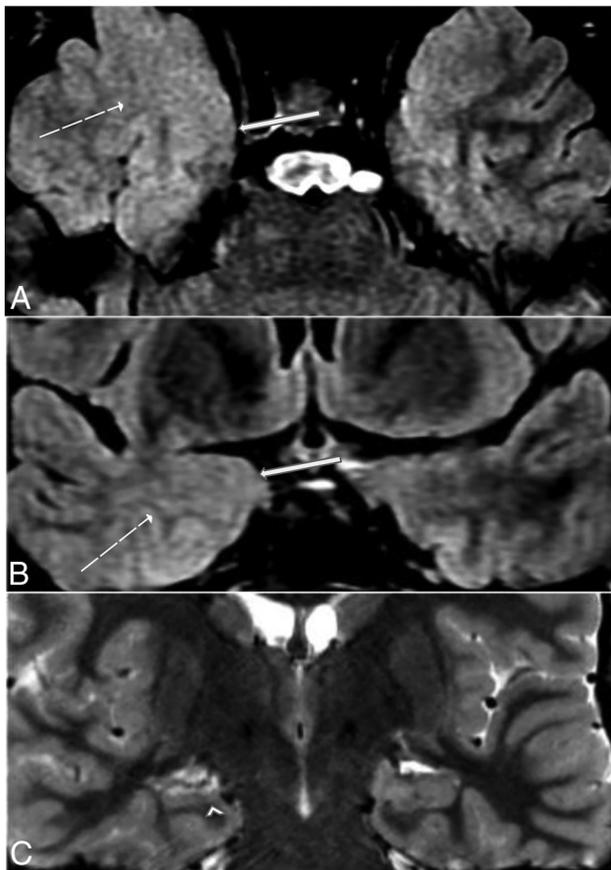
Multiple factors including MRI protocol, magnet strength, reader experience, and correlation with MEG or EEG findings are crucial in diagnosing FCD.<sup>34-36</sup> Ultra-high-field strength 7T may aid in diagnosis of FCD, including in MRI-negative patients.<sup>34,35</sup> Postprocessing by using morphometric analysis has shown promise in detecting subtle lesions.<sup>34,37</sup>

A large study by Wang et al<sup>38</sup> reported that FCD type I comprised most FCD (37/43) in 95 MRI-negative surgical patients. It could be surmised that while FCD was missed in 27 of 56 total FCD cases in our cohort of 886 patients, the number may be larger.

FCD type Ia and Ib demonstrate subtle increased T2-FLAIR signal in the white matter, and FCD type Ia demonstrates decreased white matter volume.<sup>34</sup> Features of FCD type II on MRI include altered cortical thickness, FLAIR hyperintensity of the lesion and adjacent white matter, and blurring of the gray-white matter junction (Figs 2 and 3). FCD type IIb, which has balloon cells,



**FIG 2.** FCD: (A) sagittal FLAIR image and (B) coronal FLAIR image from a patient with intractable seizures. Left frontal onset. FLAIR images demonstrate obscuration of the gray-white matter interface involving the frontal operculum along the anterior extent of the pars triangularis (arrows).



**FIG 3.** Dual Pathology, FCD missed: patient with missed diagnosis of FCD, in the dual category. Axial FLAIR, coronal FLAIR, and coronal T2-weighted sequences, respectively. A, and B, coronal FLAIR image, demonstrate cortical thickening in the medial right temporal lobe (solid arrow) and blurring of the gray-white matter junction, and FLAIR hyper-signal in the white matter (dashed arrow). C, Coronal T2 image demonstrates loss of height and internal architecture of the right hippocampus and T2 signal hyperintensity consistent with previously diagnosed right MTS (arrowhead).

can be associated with linear or triangular shaped T2/FLAIR hyperintensity extending between the lesion and ventricle. Ten of 886 (1.1%) patients had features of FCD type IIb on MRI. No FCD of the IIb variety were missed.

Utilization of the recommended high-resolution 3D T1 and FLAIR sequences, in conjunction with clinical and EEG findings allows for optimized assessment for FCD.<sup>6,12</sup> It must be stated that FCD ranges from truly invisible on imaging to subtle to visible (FCD type IIb). We hope to highlight this as a possible missed entity.

### Encephaloceles

Encephaloceles were identified in 26 of 886 patients (2.8%), 25 of which were temporal in location. Ten of 26 (38.5%) were missed.

Encephaloceles are related to either a congenital or acquired osseous and dural defect in the skull base. The literature reports an overall incidence of approximately 1 in 35,000.<sup>39,40</sup> While they may cause a host of clinical problems such as CSF leak, meningitis, middle ear effusions, or conductive hearing loss, they cause MRE in a small percentage of patients.<sup>39</sup> Saavalainen et al<sup>40</sup> reported temporal anteroinferior encephalocele in 0.3% of patients newly diagnosed with epilepsy, with a higher frequency of 1.9% in patients with MRE at their center.

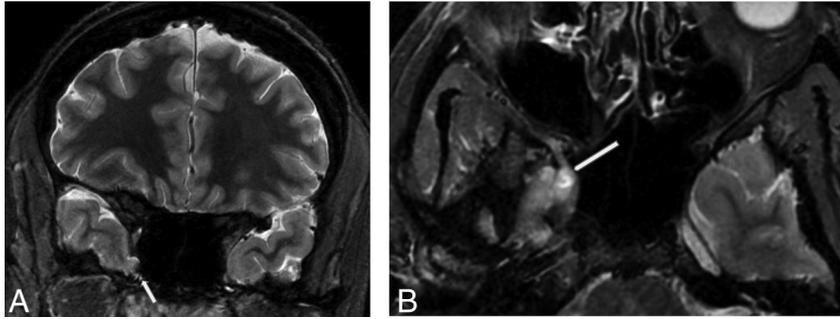
Encephaloceles, and specifically temporal encephaloceles, may be missed in routine MRI interpretation during epilepsy work-up.<sup>41</sup> Imaging findings are often subtle, and interpretation is improved if relevant clinical information, including seizure semiology and EEG reports are provided. Multiple studies emphasize temporal encephaloceles as an overlooked cause for MRE on MRI-negative studies.<sup>18,40</sup>

MRI features include distortion of the skull base contour, associated bulging of brain parenchyma through this region, surrounded by a rim of CSF herniating through the defect (Fig. 4).<sup>18</sup> Traction on herniated brain is thought to result in ischemia and gliosis.<sup>42</sup> It is imperative to look closely for these features when interpreting any epilepsy MRI study.

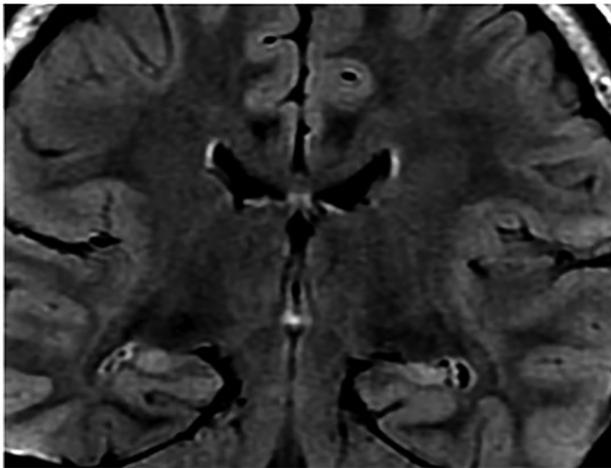
In their study, Toledano et al<sup>18</sup> reported an average delay of 9 years between onset of epilepsy and diagnosis of temporal encephalocele and highlighted the importance of specifically searching for these lesions in patients with MRE-TLE. Campbell et al<sup>43</sup> reported an increased detection rate in those who displayed coexisting MRI signs of idiopathic intracranial hypertension. Morone et al<sup>41</sup> described superior detection of temporal encephalocele on 7T MRI, however, these are not widely available for clinical use.

### MTS

MTS was the most common positive diagnosis overall in our study. It is well-documented in TLE and may progress to MRE in up to 90% of patients.<sup>34</sup> MTS was seen in 266 patients (30%), of whom 233 (87.6%) were unilateral, and 33 (12.4%) were bilateral. In 71 (26.7%), MTS was part of dual pathology. A relatively higher proportion of bilateral MTS was missed compared with unilateral MTS on initial read. This was presumed in part attributable to perceived symmetry of the abnormality in cases of bilateral MTS. Six



**FIG 4.** Encephalocele: A, coronal T2 and B, axial T2 images from a patient with temporal lobe seizures, EEG localizing to the right frontotemporal region with evidence of EA. MRI demonstrates a very small right temporopolar medially oriented encephalocele with associated abnormal signal at the level of its exit point.



**FIG 5.** Bilateral MTS: coronal FLAIR image demonstrates decreased height, FLAIR hyperintensity, and loss of internal architecture in the bilateral hippocampi consistent with bilateral MTS, worse on the left. This was missed on the right on initial interpretations.

of 33 (18.2%) cases of bilateral MTS were originally misdiagnosed as unilateral on initial read due to asymmetrical severity of the imaging findings (Fig 5). MTS was missed in 15 patients in whom an additional alternative diagnosis was identified and reported.

In MTS, MRI demonstrates atrophy, loss of internal architecture, increased T2 signal and decreased T1 signal of the involved hippocampus. Findings can be subtle showing only loss of internal architecture. Up to 29% of surgically proved MTS was previously reported as MRI-negative.<sup>10,38,44</sup>

It is imperative to analyze the hippocampi carefully to ensure signal, volume, and internal architecture are maintained. Extrahippocampal abnormalities may assist in ultimately detecting findings of MTS, including atrophy or T2 hyperintense signal of ipsilateral amygdala, temporal white matter, temporal neocortex, fornix, mammillary bodies, thalamus, insula or basal frontal cortex.

### Others

Ulegyria was present in 18 of 886 (2%) patients and 7 (38.9%) ulegyria diagnoses were missed. Ulegyria develops because of a perinatal hypoxic-ischemic brain injury in term neonates. It is said to confer a mushroom-shaped configuration to involved

gyri.<sup>45</sup> While rare, it is considered highly epileptogenic. As ulegyria results from a perinatal insult, pediatric neuro-radiologists may be more familiar with it. The relative infrequency of ulegyria in adults may account for this having been missed in a small number of our cohort.

Fifty of 886 patients had a radiologic diagnosis of tumor. Small tumors were missed in 4 of 50 (8%) including 2 low-grade gliomas, 1 dysembryoplastic neuroepithelial tumor, and 1 hypothalamic hamartoma in a patient posttemporal lobectomy.

Many additional pathologies, including rare entities, were noted in our patient cohort. These are included only in the Supplemental Data.

In general, to maximize interpretation of epilepsy-MRI and increase detection rate for subtle abnormalities, a range of imaging-specific tools and techniques can be utilized. This includes high-resolution MRI protocols including dedicated specialized epilepsy MRI protocols on a minimum field strength of 3T, which offer better visualization of structural abnormalities.<sup>12</sup> The 3D T1-weighted sequence is optimized for anatomy, 3D-FLAIR sequence for signal abnormalities and high in-plane resolution, and 2D coronal T2-weighted sequence for evaluation of the hippocampal structures.<sup>12</sup> Specifically, narrowing the window width on FLAIR images can accentuate gliosis associated with MTS, while similar adjustments on T2-weighted images can highlight the subcortical hyperintensities characteristic of FCD.<sup>46</sup> Multiplanar analysis of the calvaria and skull base is imperative to optimize detection of small encephaloceles.<sup>43-45</sup>

Utilization of advanced imaging modalities such as diffusion tensor imaging and functional MRI provides a comprehensive view of both functional and structural brain abnormalities in patients with epilepsy, enhancing the accuracy of lesion detection and improving surgical outcomes.<sup>17</sup> PET-MRI and subtraction ictal SPECT coregistered to MRI can also provide invaluable insights in localizing the epileptogenic zone, especially when MRI findings are inconclusive.<sup>47,48</sup>

Additional tools including those not currently used in our institution, can be helpful, including but not limited to: volumetric analysis that can aid in the assessment of hippocampal volume, detect subtle MTS or region-specific volume loss,<sup>32,34</sup> voxel-based morphometry; or advanced postprocessing software tools,<sup>37</sup> which can assist in assessment of surface anatomy, tissue contrast, and FCD. MR spectroscopy can be helpful in certain cases, such as those in which an underlying metabolic condition is suspected. Research has also demonstrated that MR spectroscopy can detect metabolic abnormalities in the epileptogenic zone, when structural MRI appeared normal, potentially aiding the presurgical evaluation of patients with epilepsy.<sup>49</sup>

A standardized approach to epilepsy MRI interpretation, including the use of template reporting may be beneficial in improving accuracy, especially with time constraints in mind.<sup>50</sup> Common miss areas, such as those highlighted in this article should be specifically sought.

The need for interdisciplinary clinical information is clear. As a rule, this should be included where possible in the MRI requisition, including seizure semiology and EEG findings when available. Similarly, the radiologist should be cognizant of this and seek this information where possible, including use of the EMR.

With future directions in mind, we could consider standardized reporting templates, postprocessing, and artificial intelligence techniques that may further aid the radiologist in subtle lesion detection.

### Limitations

Limitations of this study include the retrospective nature of the study. This relies on pre-existing data, which may result in selection bias, because the cohort was restricted to patients discussed at a single tertiary epilepsy center. Furthermore, the single-center nature of the study may limit the generalizability of findings to institutions with different imaging protocols or patient populations. No pediatric patients were included. This reflects normal practice in many centers, where adult and pediatric patients are assessed separately. Abnormalities of cortical development are more common in the pediatric setting, and thus our numbers of cortical and migrational abnormalities are relatively low. This was deemed acceptable as this reflects normal practice for adult neuroradiologists. The single-center nature also impacts prevalence of more location-specific pathologies, which may be endemic to certain regions globally. The lack of histopathologic confirmation represents a limitation.

Our study relied on consensus reads during multidisciplinary rounds as the final diagnosis, which, while thorough, introduces the possibility of confirmation bias, as additional clinical and EEG information may have influenced interpretations. A single subspecialized neuroradiologist identified all discrepancies, which could overemphasize the role of expertise in detecting subtle findings and underestimate the performance of less experienced radiologists. Last, because this study analyzed reports retrospectively, it is difficult to quantify how external factors, such as time pressures or workload, contributed to misinterpretations.

### CONCLUSIONS

This study emphasizes the importance of interdisciplinary collaboration in accurately interpreting MRI for MRE. Despite optimized imaging protocols, 16.7% of initial MRI interpretations were discrepant, highlighting the necessity to combine imaging, clinical, and EEG data. Discrepancies were not linked to radiologists' experience, underscoring the need for ongoing training. Better integration of clinical data into reporting workflows could enhance diagnostic accuracy and outcomes, as might incorporation of tools such as artificial intelligence.

We hope to inform radiologists of commonly overlooked pathologies in brain MRI interpretation for patients with epilepsy to maximize the diagnostic yield of initial MRI interpretation in these patients.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

### REFERENCES

1. Beghi E. **The epidemiology of epilepsy.** *Neuroepidemiology* 2020; 54:185–91 CrossRef Medline
2. Fisher RS, Boas WE, Blume W, et al. **Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE).** *Epilepsia* 2005;46:470–72 CrossRef Medline
3. Ngugi AK, Kariuki SM, Bottomley C, et al. **Incidence of epilepsy: a systematic review and meta-analysis.** *Neurology* 2011;77:1005–12 CrossRef Medline
4. Fiest KM, Sauro KM, Wiebe S, et al. **Prevalence and incidence of epilepsy: a systematic review and meta-analysis of international studies.** *Neurology* 2017;88:296–303 CrossRef Medline
5. Beghi E, Giussani G, Sander JW. **The natural history and prognosis of epilepsy.** *Epileptic Disord* 2015;17:243–53 CrossRef Medline
6. Von Oertzen J, Urbach H, Jungbluth S, et al. **Standard magnetic resonance imaging is inadequate for patients with refractory focal epilepsy.** *J Neurol Neurosurg Psychiatry* 2002;73:643–47 CrossRef Medline
7. Shorvon SD, Goodridge DM. **Longitudinal cohort studies of the prognosis of epilepsy: contribution of the National General Practice Study of Epilepsy and other studies.** *Brain* 2013;136:3497–510 CrossRef Medline
8. López González FJ, Rodríguez Osario X, Gil-Nagel Rein A, et al. **Drug-resistant epilepsy: definition and treatment alternatives.** *Neurologia* 2015;30:439–46
9. Kwan P, Brodie MJ. **Early identification of refractory epilepsy.** *N Engl J Med* 2000;342:314–19 CrossRef Medline
10. Hainc N, McAndrews MP, Valiante T, et al. **Imaging in medically refractory epilepsy at 3 Tesla: a 13-year tertiary adult epilepsy center experience.** *Insights Imaging* 2022;13:99 CrossRef Medline
11. Bernasconi A, Cendes F, Theodore WH, et al. **Recommendations for the use of structural magnetic resonance imaging in the care of patients with epilepsy: a consensus report from the International League Against Epilepsy Neuroimaging Task Force.** *Epilepsia* 2019; 60:1054–68 CrossRef Medline
12. Blumcke I, Spreafico R, Haaker G; EEBB Consortium, et al. **Histopathological findings in brain tissue obtained during epilepsy surgery.** *N Engl J Med* 2017;377:1648–56 CrossRef Medline
13. Cankurtaran CZ, Templer J, Bandt SK, et al. **Multimodal presurgical evaluation of medically refractory focal epilepsy in adults: an update for radiologists.** *AJR Am J Roentgenol* 2022;219:488–500 CrossRef Medline
14. Urbach H, Hattingen J, von Oertzen J, et al. **MR imaging in the presurgical workup of patients with drug-resistant epilepsy.** *AJNR Am J Neuroradiol* 2004;25:919–26 Medline
15. McGonigal A, Bartolomei F, Régis J, et al. **Stereoencephalography in presurgical assessment of MRI-negative epilepsy.** *Brain* 2007;130: 3169–83 CrossRef Medline
16. Muhlhofer W, Tan YL, Mueller SG, et al. **MRI-negative temporal lobe epilepsy—what do we know?** *Epilepsia* 2017;58:727–42 CrossRef Medline
17. Cascino GD, Jack CR, Jr, Parisi JE, et al. **Magnetic resonance imaging-based volume studies in temporal lobe epilepsy: pathological correlations.** *Ann Neurol* 1991;30:31–36 CrossRef Medline
18. Toledano R, Garcia-Morales I, Jimenez-Huete A, et al. **Small temporal pole encephalocoele: a hidden cause of “normal” MRI temporal lobe epilepsy.** *Epilepsia* 2016;57:841–51 CrossRef Medline
19. Duncan JS. **Imaging in the surgical treatment of epilepsy.** *Nat Rev Neurol* 2010;6:537–50 CrossRef Medline
20. Bien CG, Szinay M, Wagner J, et al. **Characteristics and surgical outcomes of patients with refractory magnetic resonance imaging-negative epilepsies.** *Arch Neurol* 2009;66:1491–99 CrossRef Medline
21. Nguyen DK, Mbacou MT, Nguyen DB, et al. **Prevalence of nonlesional focal epilepsy in an adult epilepsy clinic.** *Can J Neurol Sci* 2013;40:198–202 CrossRef Medline
22. Makhalova J, Le Troter A, Aubert-Conil S, et al. **Epileptogenic networks in drug-resistant epilepsy with amygdala enlargement:**

- assessment with stereo-EEG and 7T MRI. *Clin Neurophysiol* 2022;133:94–103 CrossRef Medline
23. Tebartz van Elst L, Woermann FG, Lemieux L, et al. **Amygdala enlargement in dysthymia—a volumetric study of patients with temporal lobe epilepsy.** *Biol Psychiatry* 1999;46:1614–23 CrossRef Medline
  24. Beh SMJ, Cook MJ, D'Souza WJ. **Isolated amygdala enlargement in temporal lobe epilepsy: a systematic review.** *Epilepsy Behav* 2016;60:33–41 CrossRef Medline
  25. Coan AC, Morita ME, de Campos BM, et al. **Amygdala enlargement in patients with mesial temporal lobe epilepsy without hippocampal sclerosis.** *Front Neurol* 2013;4:166 CrossRef Medline
  26. Reyes A, Thesen T, Kuzniecky R, et al. **Amygdala enlargement: temporal lobe epilepsy subtype or nonspecific finding?** *Epilepsy Res* 2017;132:34–40 CrossRef Medline
  27. Suzuki H, Sugano H, Nakajima M, et al. **The epileptogenic zone in pharmaco-resistant temporal lobe epilepsy with amygdala enlargement.** *Epileptic Disord* 2019;21:252–64 CrossRef Medline
  28. Bower SPC, Vogrin SJ, Morris K, et al. **Amygdala volumetry in “imaging-negative” temporal lobe epilepsy.** *J Neurol Neurosurg Psychiatry* 2003;74:1245–49 CrossRef Medline
  29. Van Paesschen W, Connelly A, Johnson CL, et al. **The amygdala and intractable temporal lobe epilepsy: a quantitative magnetic resonance imaging study.** *Neurology* 1996;47:1021–31 CrossRef Medline
  30. Na HK, Lee H-J, Hong S-J, et al. **Volume change in amygdala enlargement as a prognostic factor in patients with temporal lobe epilepsy: a longitudinal study.** *Epilepsia* 2020;61:70–80 CrossRef Medline
  31. Malter MP, Widman G, Galldikis N, et al. **Suspected new-onset autoimmune temporal lobe epilepsy with amygdala enlargement.** *Epilepsia* 2016;57:1485–94 CrossRef Medline
  32. Shakhatreh L, Sinclair B, McLean C, et al. **Amygdala enlargement in temporal lobe epilepsy: histopathology and surgical outcomes.** *Epilepsia* 2024;65:1709–19 CrossRef Medline
  33. Aminpour A, Ebrahimi M, Widjaja E. **Lesion segmentation in paediatric epilepsy utilizing deep learning approaches.** *AAIML* 2022;02:422–40 CrossRef
  34. Najm I, Lal D, Vanegas MA, et al. **The ILAE consensus classification of focal cortical dysplasia: an update proposed by an ad hoc task force of the ILAE Diagnostic Methods Commission.** *Epilepsia* 2022;63:1899–919 CrossRef Medline
  35. Wang I, Bernasconi A, Bernhardt B, et al. **MRI essentials in epilepsy: a review from the ILAE Imaging Taskforce.** *Epileptic Disord* 2020;22:421–37 CrossRef Medline
  36. Wang I, Oh S, Blümcke I, et al. **Value of 7T MRI and post-processing in patients with nonlesional 3T MRI undergoing epilepsy presurgical evaluation.** *Epilepsia* 2020;61:2509–20 CrossRef Medline
  37. Wang ZI, Alexopoulos AV, Jones SE, et al. **Linking MRI postprocessing with magnetic source imaging in MRI-negative epilepsy.** *Ann Neurol* 2014;75:759–70 CrossRef Medline
  38. Wang ZI, Alexopoulos AV, Jones SE, et al. **The pathology of magnetic-resonance-imaging-negative epilepsy.** *Mod Pathol* 2013;26:1051–58 CrossRef Medline
  39. Wind JJ, Caputy AJ, Roberti F. **Spontaneous encephalocoeles of the temporal lobe.** *Neurosurg Focus* 2008;25:E11 CrossRef Medline
  40. Saavalainen T, Jutila L, Mervaala E, et al. **Temporal anteroinferior encephalocoele: an underrecognized etiology of temporal lobe epilepsy?** *Neurology* 2015;85:1467–74 CrossRef Medline
  41. Morone PJ, Sweeney AD, Carlson ML, et al. **Temporal lobe encephalocoeles: a potentially curable cause of seizures.** *Otol Neurotol* 2015;36:1439–42 CrossRef Medline
  42. Faulkner HJ, Sandeman DR, Love S, et al. **Epilepsy surgery for refractory epilepsy due to encephalocoele: a case report and review of the literature.** *Epileptic Disord* 2010;12:160–66 CrossRef Medline
  43. Campbell ZM, Hyer JM, Lauzon S, et al. **Detection and characteristics of temporal encephalocoeles in patients with refractory epilepsy.** *AJNR Am J Neuroradiol* 2018;39:1468–72 CrossRef Medline
  44. Smith AP, Sani S, Kanner AM, et al. **Medically intractable temporal lobe epilepsy in patients with normal MRI: surgical outcome in twenty-one consecutive patients.** *Seizure* 2011;20:475–79 CrossRef Medline
  45. Bican Demir A, Eser P, Bekar A, et al. **A rare cause of epilepsy: ulegyria revisited in a series of 10 patients.** *Clin EEG Neurosci* 2022;53:153–59 CrossRef Medline
  46. Ho CY, Radhakrishnan R. **Optimal epilepsy imaging and protocols.** *Adv Clin Radiol* 2021;3:253–63 CrossRef
  47. Bouilleret V, Valenti MP, Hirsch E, et al. **Correlation between PET and SISCOM in temporal lobe epilepsy.** *J Nucl Med* 2002;43:991–98 Medline
  48. Traub-Weidinger T, Arbizu J, Barthel H, et al. **EANM practice guidelines for an appropriate use of PET and SPECT for patients with epilepsy.** *Eur J Nucl Med Mol Imaging* 2024;51:1891–908 CrossRef Medline
  49. Caruso PA, Johnson J, Thibert R, et al. **The use of magnetic resonance spectroscopy in the evaluation of epilepsy.** *Neuroimaging Clin N Am* 2013;23:407–24 CrossRef Medline
  50. Nobel JM, van Geel K, Robben SGF. **Structured reporting in radiology: a systematic review to explore its potential.** *Eur Radiol* 2022;32:2837–54 CrossRef Medline

# Cross-Sectional Validation of an Automated Lesion Segmentation Software in Multiple Sclerosis: Comparison with Radiologist Assessments

 Maria Vittoria Spampinato,  Heather R. Collins, Hannah Wells, William Dennis,  Jordan H. Chamberlin,  Emily Ye, Justin A. Chetta,  Maria Gisele Matheus,  Seth T. Stalcup, and  Donna R. Roberts



## ABSTRACT

**BACKGROUND AND PURPOSE:** MRI is widely used to assess disease burden in MS. This study aimed to evaluate the effectiveness of a commercially available k-nearest neighbors (kNN) network software in quantifying white matter lesion (WML) burden in MS. We compared the software's WML quantification to expert radiologists' assessments.

**MATERIALS AND METHODS:** We retrospectively reviewed brain MRI examinations of adult patients with MS and of adult patients without MS and with a normal brain MRI referred from the neurology clinic. MR images were processed by using an AI-powered, cloud-based kNN software, which generated a DICOM lesion distribution map and a report of WML count and volume in 4 brain regions (periventricular, deep, juxtacortical, and infratentorial white matter). Two blinded radiologists performed semiquantitative assessments of WM lesion load and lesion segmentation accuracy. Additionally, 4 blinded neuroradiologists independently reviewed the data to determine if MRI findings supported an MS diagnosis. The associations between radiologist-rated WML load and kNN model WML volume and count were evaluated with Spearman rank order correlation coefficient ( $\rho$ ) because these variables were not normally distributed. Results were considered significant when  $P < .05$ .

**RESULTS:** The study included 32 patients with MS (35.4 years  $\pm$ 9.1) and 19 patients without MS (33.5 years  $\pm$ 12.1). The kNN software demonstrated 94.1% and 84.3% accuracy in differentiating MS from non-MS subjects based respectively on WML count and WML volume, compared with radiologists' accuracy of 90.2% to 94.1%. Lesion segmentation was more accurate for the deep WM and infratentorial regions than for the juxtacortical region (both  $P < .001$ ).

**CONCLUSIONS:** kNN-derived WML volume and WML count provide valuable quantitative metrics of disease burden in MS. AI-powered postprocessing software may enhance the interpretation of brain MRIs in MS patients.

**ABBREVIATIONS:** 3D DIR = 3D double-inversion recovery; AUC = area under the curve; EDSS = Expanded Disability Status Scale; FN = false-negative; FP = false-positive; ICC = intraclass correlation coefficient; kNN = k-nearest neighbors; MP2RAGE = magnetization prepared 2 rapid acquisition gradient echoes; ROC = receiver operating characteristic; SPACE = sampling perfection with application-optimized contrasts by using a different flip angle evolution; WMH = white matter hyperintensity; WML = white matter lesion

MS is an autoimmune disorder of the central nervous system characterized by recurrent episodes of inflammation, demyelination, and axonal loss, which can lead to a progressive accumulation of neurologic disability. The McDonald criteria for diagnosing MS rely on clinical evaluation, laboratory tests, and presence of lesions on T2-weighted images, particularly FLAIR

MRI sequences.<sup>1</sup> MRI supports the diagnosis by demonstrating whether the criteria for dissemination in time and space are met.<sup>1,2</sup> Lesion segmentation can provide imaging biomarkers of disease burden.<sup>3</sup> However, manual delineation of MRI lesions requires considerable time, resources, and technical expertise.

Automated lesion segmentation methods offer objectivity and speed for lesion detection and quantification and have the potential to improve care of patients with MS.<sup>4</sup> Artificial intelligence offers powerful solutions for medical image postprocessing. Machine learning can automate complex tasks, such as image segmentation, and perform these analyses fast with high accuracy and reproducibility.<sup>5</sup> k-nearest neighbors (kNN) networks, a machine learning algorithm that classifies data points by analyzing the "k" most similar neighbors in space, have been employed for lesion segmentation in MS and other disorders.<sup>6-10</sup> Previous studies evaluating automated white matter lesion (WML) segmentation

Received September 6, 2024; accepted after revision December 30.

From the Department of Radiology (M.V.S., H.R.C., W.D., J.H.C., J.A.C., M.G.M., S.T.S., D.R.R.), and College of Medicine (H.W., E.Y.), Medical University of South Carolina, Charleston, South Carolina.

This research has been supported by a grant from Siemens Healthineers.

Please address correspondence to Maria V. Spampinato, MD, Department of Radiology and Radiological Science, Medical University of South Carolina, 96 Jonathan Lucas St, MSC 323, Charleston, SC 29425; e-mail: spampin@musc.edu; @mvspampinato, @MUSCNeurorads



Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8655>

## SUMMARY

**PREVIOUS LITERATURE:** Previous studies evaluating automated WML segmentation algorithms, including kNN, primarily used 3T MRI data sets and focused on overlap-based metrics like the Dice similarity coefficient. Few studies explored the algorithm's real-world performance across varied MRI field strengths or compared outputs directly to radiologist assessments, an important factor for clinical applicability.

**KEY FINDINGS:** The kNN algorithm demonstrated high accuracy in WML segmentation, particularly in infratentorial and deep white matter. The kNN software demonstrated comparable accuracy to radiologists in differentiating MS from non-MS subjects, with better performance of kNN-based WM lesion count than WM lesion volume.

**KNOWLEDGE ADVANCEMENT:** This study evaluates the use of a kNN-based WML segmentation algorithm in an external cohort and highlights its potential to streamline radiology workflows by providing objective WML quantification while identifying areas for refinement, particularly in juxtacortical lesion segmentation.

by using kNN algorithms, primarily used 3T MRI data sets and focused on overlap-based metrics like the Dice similarity coefficient. Few studies explored the algorithm's real-world performance across varied MRI field strengths or compared outputs directly to radiologist assessments, an important factor for clinical applicability.

Our purpose was to evaluate the performance of a commercially available kNN algorithm in quantifying WML burden in brain MRI examinations of patients with MS, compared with expert radiologists, by using an external validation data set.

## MATERIALS AND METHODS

### Patient Cohort

This article follows the STARD reporting guidelines. Institutional review board approval was obtained for this retrospective single-center cross-sectional validation study, including a waiver of informed consent. We retrospectively reviewed all brain MRI examinations performed at our hospital system between July and September 2021 by using a newly implemented MRI protocol for MS (described below) during the first 3 months of utilization of the protocol. Inclusion criteria were the following: 1) age range = 18–65 years; 2) MS diagnosis<sup>1</sup>; 3) brain MRI performed per our institution's MS protocol, including a 3D T2 FLAIR sampling perfection with application-optimized contrasts by using a different flip angle evolution (SPACE) sequence<sup>11</sup> and 3D MPRAGE without intravenous contrast administration; and 4) neurologic evaluation within 1 month of the MRI. We excluded patients with neurologic disorders other than MS, incomplete studies, or MRI examinations degraded by head motion or other artifacts. The degree of disability was assessed by using the Kurtzke Expanded Disability Status Scale (EDSS),<sup>12</sup> ranging from 0 to 10, with higher scores meaning more severe disability. As a comparison group, we included consecutive patients without MS who were referred to our imaging facilities from the neurology clinic between July and September 2021 with the following inclusion criteria: 1) age range = 18–65 years; 2) brain MRI interpreted as normal for age by the interpreting radiologist, and 2 senior coauthors; 3) brain MR imaging including FLAIR SPACE and MPRAGE sequences; 4) no known neurologic disorder; and 5) normal neurologic evaluation within 1 month of the MR imaging.

### Imaging Protocol

MRI examinations were performed on 1.5T and 3T MR scanners (Magnetom Lumina, Skyra, Aera, Sola, and Avanto, Siemens Healthineers) by using 16- and 20-channel head coils (Supplemental Data). Preliminary image quality assessments for study inclusion were performed by 2 board-certified neuroradiologists.

### Image Postprocessing

Postprocessing was performed off-line using AI-Rad Companion Brain MR white matter hyperintensities, an AI application that extends AI-Rad Companion Brain MR brain morphometry by performing segmentation and quantification of white matter hyperintensities (WMHs). MPRAGE and FLAIR SPACE DICOM images were postprocessed by using cloud-based software. The framework combines 2 approaches: 1) A supervised method was used to obtain a map of potential lesional tissue by using a kNN classifier trained on a set of features (signal intensities, spatial location coordinates, and tissue prior probabilities) obtained from atlas-based prior probability maps of gray matter, white matter, and CSF, and 2) a Bayesian partial volume estimation algorithm.<sup>6–8</sup> The white matter hyperintensity segmentation method included the following steps (Supplemental Data): 1) preprocessing: images were aligned, skull-stripped, corrected for bias field and intensity normalized and 2) lesion segmentation was performed by a kNN classifier, in which each voxel was labeled with a value representing the probability of containing lesion or healthy tissue.<sup>13</sup> Prior probability maps were included to estimate realistic concentration maps of normal and abnormal brain tissue. Probability maps are employed to guide tissue classification (gray matter, white matter, CSF, and lesions) and to compute lesion volumes.<sup>14</sup> This algorithm automatically evaluates WMH lesion load (count and volume) while accounting for the mixing of normal and lesional tissue within voxels due to partial volume effects.<sup>7</sup>

The software generated: 1) color-coded lesion distribution maps overlaid on the sagittal FLAIR SPACE images and 2) a segmentation report of WML count (lesion number) and WML volume (lesion volume [mL]) for 4 brain subregions (periventricular, deep, juxtacortical, and infratentorial WM).

### Radiologist Imaging Interpretation

After a training session, 2 blinded board-certified radiologists, a neuroradiology attending and a neuroradiology fellow (respectively

**Table 1: Study population**

	Multiple Sclerosis	Non-Multiple Sclerosis
Number of subjects	32	19
Sex (M/F)	5/27	7/12
Age (yr) <sup>a</sup>	35.4 ± 9.1	33.5 ± 12.1
3T / 1.5T MRI	22 / 10	19 / 0
Subtype of MS	<sup>b</sup> RR/PP/SP/unknown 29/0/0/3	
Study indications	New onset of seizures (15 cases) Syncope/loss of consciousness (5 cases) First episode of psychosis (1 case) Remote meningitis (1 case)	
Years since initial MS diagnosis (yr) <sup>a</sup>	6.8 ± 5.8	N/A
EDSS <sup>c</sup>	1.3, 0.5–3	N/A
kNN total WM lesion volume <sup>a</sup>	11.0 ± 12.2	2.0 ± 3.8
kNN total WM lesion count <sup>a</sup>	36.3 ± 26.8	4.8 ± 2.7

<sup>a</sup> Mean ± standard deviation.

<sup>b</sup> Subtypes of MS: Relapsing-remitting (RR), primary-progressive (PP), secondary-progressive (SP), and unknown subtype.

<sup>c</sup> Median, interquartile range.

20 years and 4 years of experience) independently reviewed MPRAGE, FLAIR SPACE, and color-coded lesion maps superimposed on sagittal FLAIR SPACE images on Syngo.via (Siemens Healthineers) in a randomized manner. During 2 separate sessions, the neuroradiologists performed the following assessments:

**1. Radiologist-Assessed Lesion Load.** Neuroradiologists independently evaluated multiplanar 3D FLAIR images of each subject in a random order and rated lesion load of each brain region (periventricular, deep, juxtacortical, and infratentorial WM) on a scale ranging from 0 to 3 (0 = none, 1 = mild, 2 = moderate, 3 = severe) (Supplemental Data). From these ratings, we calculated the radiologist-assessed WML load as the sum of the lesion load for each region.

**2. Radiologist Assessment of Color-Coded Lesion Distribution Maps.** In a separate session, neuroradiologists evaluated the software performance by reviewing the native FLAIR images along with the model's color-coded lesion distribution maps, which were superimposed on the sagittal FLAIR SPACE images. Reader discrepancies were adjudicated during a joint reading session. Radiologists performed the following assessments.

**a. Semiquantitative assessment of false-positive and false-negative lesions by brain subregion.** Neuroradiologists provided a semiquantitative rating of the number of false-positive (FP) and false-negative (FN) lesions on the kNN-based color-coded lesion distribution map by using the following scale: 0 = no FP/FN, 1 = 1–5 FP/FN lesions, 2 = 6–10 FP/FN lesions, 3 = at least 11 FP/FN lesions.

**b. Semiquantitative evaluation of lesion size by brain subregion.** Neuroradiologists assessed how well the borders of the segmented color-coded lesions matched the lesion borders on the native FLAIR images: 1 = underestimation of lesion size, 2 = slight underestimation, 3 = accurate, 4 = slight overestimation, and 5 = overestimation.

**c. Semiquantitative ratings of overall WMH segmentation accuracy by brain subregion.** Neuroradiologists rated the overall WMH segmentation accuracy on a scale of 1 to 5: 1 = very poor, 2 = poor, 3 = fair, 4 = good, and 5 = excellent.

**3. Radiologist Diagnosis of MS.** Three blinded board-certified neuroradiologists and a neuroradiology fellow (20, 15, 10, and 4 years of experience), not involved in the previous evaluations, reviewed multiplanar FLAIR SPACE images in a randomized order and determined whether findings supported a diagnosis of MS.

### Statistical Analysis

Age, disease duration, lesion volume, and lesion count were summarized by using means and standard deviations. Independent Mann-Whitney *U* tests were used to compare MS and non-MS on the kNN WML volume and count. Sensitivity, specificity, and receiver operating characteristic (ROC) curves for the WML volume and count, with area under the curve (AUC) and 95% CI, were also reported. Radiologist-assessed WML load ratings were summarized by using medians, median absolute deviations, and ranges. Comparisons across brain regions for FP lesions, FN lesions, lesion size accuracy, and overall WML segmentation accuracy were performed by using Friedman 2-way analysis of variance by ranks, with Bonferroni correction. These data were summarized with medians and median absolute deviations. The associations between radiologist-rated WML load and kNN model WML volume and count were evaluated with Spearman rank order correlation coefficient ( $\rho$ ) because these variables were not normally distributed. Reader agreement in MS diagnosis was evaluated with a 2-way random effect, single measures, absolute agreement intraclass correlation coefficient (ICC), and 95% CI. An ROC curve was constructed by using the mean confidence values of all 4 readers, with AUC and 95% CI. Two-sided *P* values were reported, the statistical significance was set at the  $\alpha < .05$  threshold. Analyses were conducted with SPSS (Version 28, IBM).

## RESULTS

### Patient Population

Thirty-two patients with MS and 19 non-MS subjects were included in the study. See Table 1 for clinical characteristics of the study subjects. EDSS ranged between 0 and 6 with a median of 2.5 (median absolute deviation = 1.25).

**1. Automated Segmentation Report and Radiologist-Assessed Lesion Load.** Lesion segmentation was carried out by using a commercially available tool, with guidance from priori probability

**Table 2: kNN segmentation report: white matter lesion count and white matter lesion volume**

	Multiple Sclerosis		Non-Multiple Sclerosis				
	n	Mdn (MAD) <sup>a</sup>	N	Mdn (MAD)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
<b>WML count</b>							
Periventricular WM <sup>a</sup>	32	7.0 (2.0)	19	3.0 (1.0)	81.3% (63.6%–92.8%)	78.9% (54.4%–94.0%)	80.4% (66.9%–90.2%)
Juxtacortical WM	32	7.5 (4.5)	19	1.0 (1.0)	87.5% (71.0%–96.5%)	73.7% (48.8%–90.9%)	82.4% (69.1%–91.6%)
Infratentorial WM	32	1.0 (1.0)	19	0.0 (0.0)	59.4% (40.6%–76.3%)	89.5% (66.9%–98.7%)	70.6% (56.2%–82.5%)
Deep WM	32	9.5 (3.5)	19	0.0 (0.0)	96.9% (83.8%–99.9%)	94.7% (74.0%–99.9%)	96.1% (86.5%–99.5%)
Total	32	26.0 (9.0)	19	4.0 (2.0)	96.9% (83.8%–99.9%)	89.5% (66.9%–98.7%)	94.1% (83.8%–98.8%)
<b>WML volume (mL)</b>							
Periventricular WM	32	3.7 (2.4)	19	1.7 (1.6)	81.3% (63.6%–92.8%)	78.9% (54.4%–94.0%)	80.4% (66.9%–90.2%)
Juxtacortical WM	32	0.7 (0.4)	19	0.2 (0.3)	84.4% (67.2%–94.7%)	63.2% (38.4%–83.7%)	76.5% (62.5%–87.2%)
Infratentorial WM	32	0.0 (0.0)	19	0.1 (0.2)	100.0% (89.1%–100.0%)	0.0% (0.0%–17.7%)	62.7% (48.1%–75.9%)
Deep WM	32	0.1 (0.1)	19	0.0 (0.0)	96.9% (83.8%–99.9%)	94.7% (74.0%–99.9%)	96.1% (86.5%–99.5%)
Total	32	6.2 (5.9)	19	2.0 (1.8)	81.3% (63.6%–92.9%)	89.5% (66.9%–98.7%)	84.3% (71.4%–93.0%)

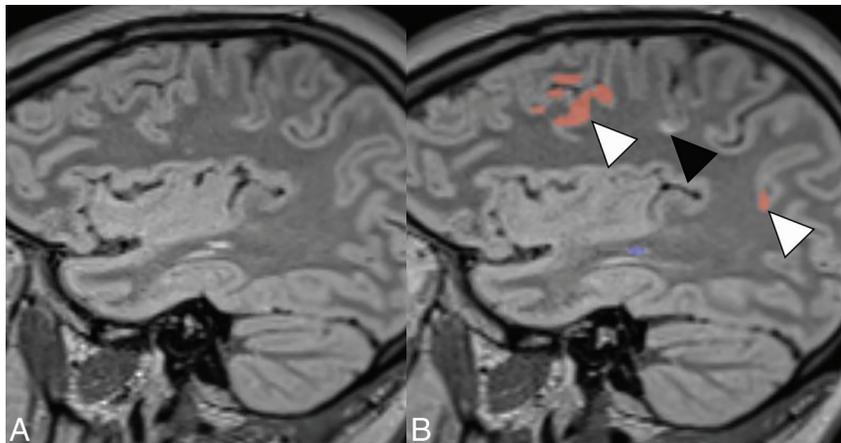
<sup>a</sup> Medians (median absolute deviations).

**Table 3: Radiologist evaluation of white matter lesion load<sup>a</sup>**

Area	Multiple Sclerosis		Non-Multiple Sclerosis	
	n	Mdn (MAD, Range) <sup>b</sup>	N	Mdn (MAD)
Periventricular	32	1 (0.14, 1–3)	19	0.5 (0.64, 0–1)
Deep WM	32	1 (0.25, 0.5–1.5)	19	0 (0.75, 0–1)
Juxtacortical	32	1 (0.29, 0–2)	19	0 (0.71, 0–1)
Infratentorial	32	1 (0.49, 0–2)	19	0 (0.51, 0–1)

<sup>a</sup> Two neuroradiologists rated WML load of each brain region on a scale ranging from 0 to 3 (lesion load, 0 = none, 1 = mild, 2 = moderate, 3 = severe).

<sup>b</sup> Medians (median absolute deviations).



**FIG 1.** Sagittal native T2 FLAIR SPACE image (A) and color-coded lesion map superimposed on the T2 FLAIR SPACE image (B) demonstrate false-positive (*white arrowhead*) and false-negative results (*black arrowhead*) for the juxtacortical region.

maps to inform tissue classification. The resulting WML counts and WML volumes (mL) for each brain subregion, and total WML counts and volumes (mL), as well as sensitivity, specificity, and accuracy, are reported in Table 2. Radiologist-assessed WML load is reported in Table 3 (Fig 4). There was no significant difference in kNN lesion count and kNN lesion volume between patients with MS who underwent brain MRI at 1.5T and those scanned at 3T, as shown by an independent-samples Mann-Whitney *U* test (lesion count:  $U = 78$ ,  $P = .192$ ; lesion volume:  $U = 64$ ,  $P = .061$ ). While lesion volume appeared slightly higher in the 1.5T subgroup, this difference was not statistically significant (Supplemental Data).

## 2. Radiologist Assessment of Color-Coded Lesion Distribution Map Accuracy.

a. Semiquantitative assessment of FP and FN lesions. In patients with MS, the juxtacortical region exhibited significantly more FP lesions than the infratentorial region ( $P < .001$ ) (Fig 1), and the periventricular region more FP lesions than the infratentorial ( $P < .001$ ) and deep WM ( $P < .001$ ). A similar pattern was observed in the non-MS group, with more FP lesions in the juxtacortical ( $P = .02$ ) and periventricular ( $P = .005$ ) than in the infratentorial regions (Table 4).

A greater number of FN lesions were found in the juxtacortical region compared with the infratentorial ( $P = .02$ ) and periventricular ( $P = .02$ ) regions in patients with MS. No FN lesions were observed in the non-MS group (Table 4).

b. Semiquantitative evaluation of lesion size by brain subregion. In the MS group, lesion size estimation was more accurate in the infratentorial ( $P = .01$ ) and deep WM regions ( $P = .006$ ) compared with the periventricular

WM. The kNN model tended to overestimate lesion size in the periventricular and juxtacortical regions (Table 4).

c. Semiquantitative ratings of overall WMH segmentation accuracy. Radiology readers found that MS lesion segmentation accuracy was significantly worse in the juxtacortical area compared with the infratentorial ( $P = .001$ ) and deep WM ( $P < .001$ ) regions. For the non-MS group, readers rated lesion segmentation accuracy as being significantly worse in the periventricular compared with the infratentorial region,  $P = .03$  (Table 4). Additionally, no significant differences were observed in lesion segmentation accuracy by region between patients with MS who underwent brain MRI at 1.5T and those scanned at 3T (Supplemental Data).

**Table 4: Radiologist evaluation of kNN software performance<sup>a</sup>**

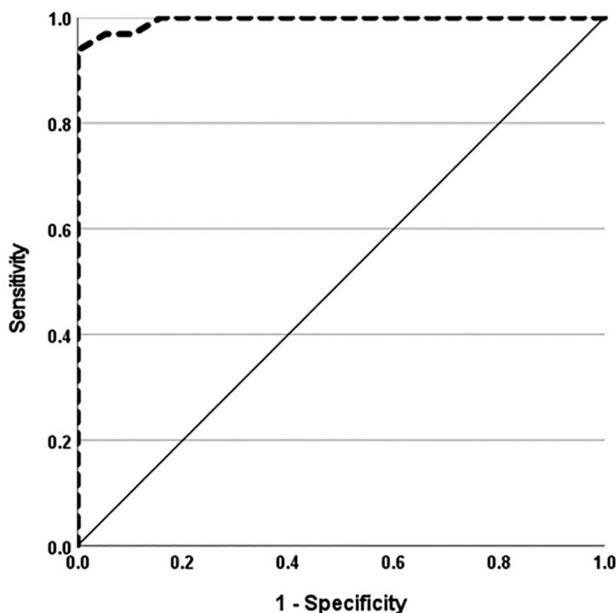
Brain Region	Multiple Sclerosis		Non-Multiple Sclerosis		All Cases	
	False-Positive <sup>b</sup>	False-Negative <sup>b</sup>	False-Positive <sup>b</sup>	False-Negative <sup>b</sup>	Lesion Size Estimation <sup>c</sup>	Overall Lesion Segmentation Accuracy <sup>d</sup>
Periventricular	1.5 (1.0)	0.5 (0.5)	1.0 (1.0)	0.0 (0.0)	3.5 (3.0–4.0)	3.5 (3.0–4.0)
Juxtacortical	1.8 (1.3)	1.0 (1.0)	1.0 (1.0)	0.0 (0.0)	4.0 (3.5–4.4)	3.0 (2.5–3.5)
Infratentorial	0.0 (0.0)	0.3 (0.3)	0.0 (0.0)	0.0 (0.0)	3.0 (3.0–3.5)	4.3 (3.0–5.0)
Deep WM	0.0 (0.0)	0.5 (0.5)	0.0 (0.0)	0.0 (0.0)	3.0 (2.8–3.5)	4.0 (3.0–5.0)

<sup>a</sup>Radiology reader ratings of FP and FN WML, summarized with medians (median absolute deviations). Radiology reader ratings for accuracy of lesion size and overall lesion segmentation accuracy, summarized with medians (interquartile ranges).

<sup>b</sup>Rating scale for false-positive/negative lesions: 0, no false-positive/negative lesions; 1, 1–5 false-positive/negative; 2, 6–10/false-positive/negative; 3, at least 11 false-positive.

<sup>c</sup>Rating scale for lesion size estimation: 1, underestimation; 2, slight underestimation; 3, accurate; 4, slight overestimation; and 5, overestimation.

<sup>d</sup>Rating scale for overall lesion segmentation accuracy: 1, very poor; 2, poor; 3, fair; 4, good; and 5, excellent.

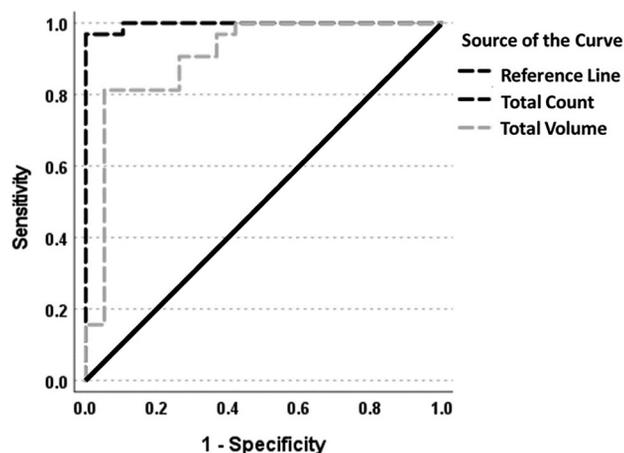


**FIG 2.** The ROC curve for all 4 readers yielded an AUC of 0.995, 95% CI = 0.921–1.000,  $P < .001$ .

**3. Radiologist Diagnosis of MS.** The 4 neuroradiologists demonstrated excellent agreement in classifying subjects as either MS or non-MS, ICC = 0.86, 95% CI = 0.79–0.91. Accuracy ranged from 90.2% to 94.1%, sensitivity ranged from 84.4% to 93.8%, and specificity ranged from 94.7% to 100.0% (Supplemental Data). The ROC curve for all 4 readers yielded an AUC of 0.995, 95% CI = 0.984–1.00,  $P < .001$  (Fig 2).

**4. Classification of MS and Non-MS Subjects Using the kNN Model Report.** The kNN-based total WML count and volumes (mL) were significantly greater in MS compared with non-MS subjects (both  $P < .001$ ). The total WML count in the deep and infratentorial WM showed respectively excellent (94.7%) and very good sensitivity (89.5%) in the classification of MS versus non-MS subjects. The sensitivity of periventricular and juxtacortical WML counts were respectively 78.9% and 73.7%. The specificity of WML count was above 80% across all brain regions (periventricular: 81.3%, juxtacortical: 87.5%, deep WM: 96.9%) (Table 2).

With respect to WML volume (mL), the sensitivity of the deep WM lesion volume was 85.7%. The sensitivity for periventricular,



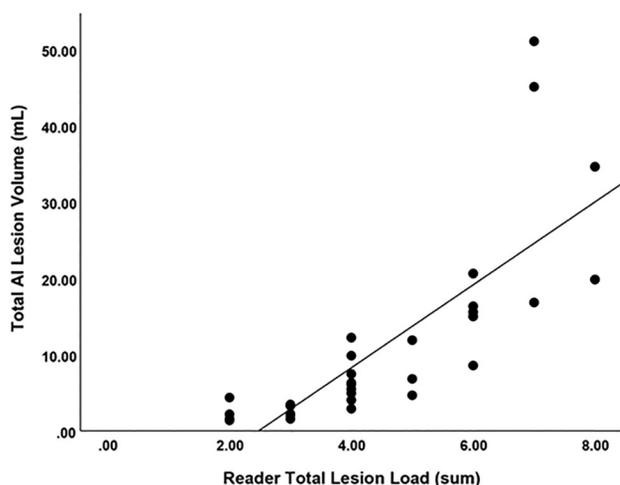
**FIG 3.** ROC curves for the total WMH count (AUC = 0.997; 95% CI = 0.924–1.000;  $P < .001$ ) and total WMH volume (AUC = 0.905; 95% CI = 0.789–0.969;  $P < .001$ ).

juxtacortical, and infratentorial WML volume were respectively 58.8%, 46.2%, and 50.0%. Specificity was above 80% for the periventricular (90.6%) and juxtacortical (90.0%) areas (Table 2). The ROC curve for WML count and volume respectively yielded an AUC of 0.997, 95% CI = 0.924–1.000,  $P < .001$ , and an AUC of 0.905, 95% CI = 0.789–0.969,  $P < .001$  (Fig 3).

## DISCUSSION

We evaluated the performance of a commercially available kNN algorithm in the segmentation of WML and in the differentiation of patients with MS from with MS non-MS patients. We found that the accuracy of lesion segmentation varied across brain regions and was most accurate for the deep WM and infratentorial regions. The kNN software demonstrated comparable accuracy with radiologists in differentiating MS from non-MS subjects, with better performance of kNN-based WM lesion count than WM lesion volume.

The kNN models have been applied to MS lesion segmentation.<sup>6,7,9,10</sup> The accuracy of kNN classification of MS WM lesions can be improved by optimizing signal intensity normalization and by adding in the model GM, WM, and CSF tissue priors from healthy controls.<sup>10</sup> Fartaria et al<sup>6</sup> tested a kNN lesion segmentation method by using 3D FLAIR, MPRAGE, magnetization-prepared 2 rapid acquisition gradient echoes (MP2RAGE), and 3D double-inversion recovery (3D DIR) in patients with MS



**FIG 4.** The scatterplot shows the kNN model total lesion volume (mL) and reader total lesion load values,  $\rho = 0.91$ ,  $P < .001$ .

with low disability status. The combination of routine clinical sequences, such as MPRAGE and 3D FLAIR, was adequate to detect small white matter lesions. However, the addition of more advanced sequences, such as MP2RAGE and 3D DIR, improved the automated detection of juxtacortical lesions. MP2RAGE and 3D DIR have relatively long acquisition times and are not routinely included in clinical MRI protocols for MS. Fartaria et al<sup>6</sup> also compared the kNN lesion segmentation method by using 3D FLAIR and MPRAGE with manual lesion segmentation and 2 open-access software applications and found that the kNN method outperformed the others because of improved segmentation of small lesions prone to partial volume effects.<sup>7</sup> The performance of commercially available kNN applications for lesion quantification in MS requires further investigation.

Our approach differed from previous studies in that we had neuroradiologists evaluate the kNN model's performance. Additionally, while most prior research was conducted by using 3T MRI,<sup>6,7,10</sup> our study included scans acquired at both 1.5T and 3T, providing a broader evaluation of the model's performance in clinical settings, where patients are commonly scanned on MRI systems at different field strengths. The kNN model segmentation accuracy varied across different regions. Specifically, the model was accurate in segmenting infratentorial and deep WM lesions, with few FP and FN WMLs reported in these regions. The evaluation of WML size was more accurate for infratentorial and deep WML than for periventricular WML. In contrast, juxtacortical and periventricular WML segmentation was less accurate, with FP WM lesions reported in both patients with MS and non-MS patients. In addition, FN WM lesions were reported for the juxtacortical region of patients with MS. Accurate segmentation of the juxtacortical and periventricular lesions is clinically important for MS diagnosis, as dissemination in the space of WMLs is a key criterion but also for monitoring disease progression.<sup>1</sup> The lower accuracy in juxtacortical and periventricular segmentation could be attributed to several factors. The proximity and the similarity in signal intensity of the juxtacortical WML and cortex, coupled with the inherent complexity and variability of gyral anatomy, may have

resulted in incorrect segmentation (see Fig 1). A potential solution would be to train the software to filter juxtacortical WMLs based on their shape, typically round, ovoid, irregular, or U-shaped, in contrast with the curvilinear cortical signal. Refining the algorithm to recognize these shapes and location characteristics could improve its ability to distinguish juxtacortical lesions from the surrounding anatomic structures. Additional training of the algorithm with a data set enriched in juxtacortical lesions might also improve segmentation accuracy in this region. Furthermore, the use of MP2RAGE and 3D DIR sequences may improve the automated detection of juxtacortical lesions.<sup>6</sup> While these sequences are relatively too long to be routinely included in clinical protocols at many institutions, faster acquisition protocols with AI-powered image reconstruction technology may soon overcome this limit.<sup>15</sup> Additionally, occurrences of incorrect periventricular segmentation were due to WML size overestimation, mislabeling of the caudate nucleus, and normal T2 FLAIR hyperintensity along the lateral ventricles. In summary, the kNN software labeled a higher number of WMLs than the radiologists, with FP results, especially in the juxtacortical and periventricular region. This issue may arise because the algorithm might be overly sensitive to certain features, leading to incorrect lesion identification. Finding a balance between sensitivity and specificity remains a key challenge in automated segmentation tools for clinical use.

The kNN model demonstrated high accuracy in differentiating MS from non-MS subjects, with an overall greater accuracy for WM lesion count (sensitivity and specificity of 89.5% and 96.9%) compared with WM lesion volume (sensitivity and specificity of 83.8% and 81.3%). Among the 4 readers, accuracy ranged from 90.2% to 94.1%, with sensitivity ranging from 84.4% to 93.8% and specificity ranging from 94.7% to 100.0%. Notably, the kNN model's performance in terms of sensitivity and specificity is comparable to the range observed among human readers, indicating its potential reliability as an adjunct tool in clinical settings.

The primary benefit of an automated WM segmentation method is its potential to streamline the radiologist interpretation workflow by providing consistent and rapid lesion quantification. Annotated FLAIR images and a report of WML count and volume would enable a rapid assessment of disease burden, potentially saving time during study interpretation. In addition, quantitative measurements could be included in the radiology report as objective markers of disease severity and progression, which, in the future, could aid clinicians in monitoring lesion burden and treatment response over time. Automated tools to improve workflow efficiency would be particularly valuable in high-volume clinical settings and for managing follow-up imaging in MS. While we have not specifically tested the utilization of the algorithm in the longitudinal assessment of MS lesion burden, this application represents an area where workflow efficiency would be highly beneficial.

Our study has several limitations, including the retrospective study design and small convenience sample size, consisting of all patients meeting the inclusion and exclusion criteria during the first 3 months of implementation of an MS MRI protocol, including 3D FLAIR and MPRAGE sequences. We acknowledge that the relatively small size of the patient and control cohorts limits the generalizability of the results. Additionally, the MRI scanners used varied across the patient and control cohorts, which

introduced variability. However, it is important to note that this algorithm has already been tested by the developers on a larger, internally controlled data set. The aim of our study was to evaluate the algorithm in an external cohort to understand its strengths and weaknesses better when applied to clinical cases where variability in imaging protocols and equipment is expected. The focus of this work is not on developing a novel method but rather on evaluating a commercially available machine learning algorithm. While deep learning segmentation methods are advancing rapidly, the purpose of this study is to assess the performance and generalizability of a commercially available machine learning tool when applied outside the original training set. Since our MS sample included predominantly mild/moderate cases (median EDSS = 2.5, range 0–6), software performance should be evaluated in a sample including patients with MS with more severe disability. Our comparison group consisted of patients referred from the neurology clinic with a brain MRI normal for age and neurologic examinations, rather than typical healthy controls. While this may introduce a bias, as this group does not represent the general population, these patients were deemed adequate as a control group for our study purposes. The kNN model produced color-coded lesion distribution maps overlaid only on sagittal FLAIR images. The reviewers noted that additional coronal and axial images would have been beneficial, especially for infratentorial assessment. We evaluated the accuracy of the model in differentiating MS from non-MS subjects as a secondary measure of the segmentation algorithm's accuracy. Our aim was to evaluate how well the algorithm's segmentation aligns with MS-related patterns identified by radiologists, not to imply that the algorithm alone could serve as a stand-alone diagnostic tool. We did not perform manual segmentation of MS lesions, nor did we evaluate the overlap between lesion masks generated by the automated algorithm and manually defined ground truth by using the Dice similarity coefficient, as has been done in previous studies assessing algorithm performance.<sup>6,7</sup> Instead, our goal was to compare the software's output with the type of semiquantitative assessments of WMH burden that are commonly employed in clinical practice. This study is structured to assess the practice utility of the software in a clinical setting where radiologists typically provide a semiquantitative assessment. We evaluated the performance of the algorithm in the detection of WMH in patients with MS, and we did not test its application in the segmentation of WMH secondary to other pathologies. As a result, our reported findings of regional performance differences are relevant only to the detection of lesion patterns in MS, not those associated with other conditions. The control group consisted of non-MS patients with brain MRI scans that were normal for age. This study design does not capture the more complex diagnostic scenarios encountered in clinical practice, such as distinguishing MS from other conditions with T2 hyperintense lesions (eg, chronic small vessel disease, other demyelinating processes). While this design allows for a controlled evaluation of the algorithm's performance, it may limit the general applicability of the results. Future studies should focus on testing the algorithm in cohorts with more diagnostically challenging differential diagnoses to assess its potential utility in clinical workflows better. In this study, we did not integrate the tool into our clinical workflow or evaluate its impact on the speed of study interpretation.

Future studies could explore how the tool might influence workflow efficiency, diagnostic accuracy, and overall clinical utility when integrated into routine practice. Despite these limitations, we believe that our study provides valuable insight into the effectiveness of the software in the evaluation of patients with MS.

## CONCLUSIONS

AI-powered postprocessing software may be a useful adjunct to the interpretation of brain MRIs in MS. The information obtained regarding cumulative WML volume and count may serve as quantitative metrics of disease burden. kNN tools for longitudinal WM lesion analyses will improve the efficiency of the radiologist workflow and serve as a valuable adjunct to radiologist interpretation.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Thompson AJ, Banwell BL, Barkhof F, et al. **Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria.** *Lancet Neurol* 2018;17:162–73 CrossRef Medline
2. Filippi M, Rocca MA. **MR imaging of multiple sclerosis.** *Radiology* 2011;259:659–81 CrossRef Medline
3. Diaz-Hurtado M, Martínez-Heras E, Solana E, et al. **Recent advances in the longitudinal segmentation of multiple sclerosis lesions on magnetic resonance imaging: a review.** *Neuroradiology* 2022;64:2103–17 CrossRef Medline
4. Egger C, Opfer R, Wang C, et al. **MRI FLAIR lesion segmentation in multiple sclerosis: does automated segmentation hold up with manual annotation?** *Neuroimage Clin* 2017;13:264–70 CrossRef Medline
5. Bonacchi R, Filippi M, Rocca MA. **Role of artificial intelligence in MS clinical practice.** *Neuroimage Clin* 2022;35:103065 CrossRef Medline
6. Fartaria MJ, Bonnier G, Roche A, et al. **Automated detection of white matter and cortical lesions in early stages of multiple sclerosis.** *J Magn Reson Imaging* 2016;43:1445–54 CrossRef Medline
7. Fartaria MJ, Todea A, Kober T, et al. **Partial volume-aware assessment of multiple sclerosis lesions.** *Neuroimage Clin* 2018;18:245–53 CrossRef Medline
8. Ayrygnac X, Le Bars E, Duflos C, et al. **Serum GFAP in multiple sclerosis: correlation with disease type and MRI markers of disease severity.** *Sci Rep* 2020;10:10923 CrossRef Medline
9. Griffanti L, Zamboni G, Khan A, et al. **BIANCA (Brain Intensity AbNormality Classification Algorithm): a new tool for automated segmentation of white matter hyperintensities.** *Neuroimage* 2016;141:191–205 CrossRef Medline
10. Steenwijk MD, Pouwels PJ, Daams M, et al. **Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs).** *Neuroimage Clin* 2013;3:462–69 CrossRef Medline
11. Wattjes MP, Ciccarelli O, Reich DS; North American Imaging in Multiple Sclerosis Cooperative MRI Guidelines Working Group, et al. **2021 MAGNIMS-CMSC-NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis.** *Lancet Neurol* 2021;20:653–70 CrossRef Medline
12. Kurtzke JF. **Rating neurologic impairment in multiple sclerosis. An expanded disability status scale (EDSS)** 1983;33:1444–1444 CrossRef
13. Wu Y, Warfield SK, Tan IL, et al. **Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI.** *Neuroimage* 2006;32:1205–15 CrossRef Medline
14. Maldjian JA, Laurienti PJ, Kraft RA, et al. **An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets.** *Neuroimage* 2003;19:1233–39 CrossRef Medline
15. Kiryu S, Akai H, Yasaka K, et al. **Clinical impact of deep learning reconstruction in MRI.** *Radiographics* 2023;43:e220133 CrossRef Medline

# Application of Deep Learning Accelerated Image Reconstruction in T2-Weighted Turbo Spin-Echo Imaging of the Brain at 7T

Zeyu Liu,  Xiangzhi Zhou,  Shengzhen Tao, Jun Ma,  Dominik Nickel, Patrick Liebig, Mahmoud Mostapha,  Vishal Patel,  Erin M. Westerhold, Hamed Mojahed, Vivek Gupta, and  Erik H. Middlebrooks



## ABSTRACT

**SUMMARY:** Prolonged imaging times and motion sensitivity at 7T necessitate advancements in image acceleration techniques. This study evaluates a 7T deep learning (DL)-based image reconstruction by using a deep neural network trained on 7T data, applied to T2-weighted turbo spin-echo imaging. Raw *k*-space data from 30 consecutive clinical 7T brain MRI patients was reconstructed by using both DL and standard methods. Qualitative assessments included overall image quality, artifacts, sharpness, structural conspicuity, and noise level, while quantitative metrics evaluated contrast-to-noise ratio (CNR) and image noise. DL-based reconstruction consistently outperformed standard methods across all qualitative metrics ( $P < .001$ ), with a mean CNR increase of 50.8% (95% CI: 43.0%–58.6%) and a mean noise reduction of 35.1% (95% CI: 32.7%–37.6%). These findings demonstrate that DL-based reconstruction at 7T significantly enhances image quality without introducing adverse effects, offering a promising tool for addressing the challenges of ultra-high-field MRI.

**ABBREVIATIONS:** CNR = contrast-to-noise ratio; DL = deep learning; GRAPPA = generalized autocalibrating partially parallel acquisitions; MNI = Montreal Neurological Institute; SAR = specific absorption rate

Ultra-high-field MRI is a potentially revolutionary technology that has been rapidly adopted after recent regulatory approvals. The impact of 7T MRI on brain imaging has already established clear advantages in diagnosis and patient care across a variety of common conditions, including epilepsy,<sup>1</sup> MS,<sup>2,3</sup> Parkinson disease,<sup>4,5</sup> stereotactic surgical planning,<sup>6</sup> functional MRI, and cerebrovascular disease.<sup>7,8</sup> Despite the advantages in contrast, SNR, and spatial resolution, there are several challenges to the routine use of clinical 7T MRI. Chief among these challenges are prolonged imaging times and increased sensitivity to motion, which necessitates advancements in acceleration of image acquisition. T2-weighted 2D TSE is a widely utilized sequence used in clinical 7T protocols due to its high in-plane resolution and favorable T2 contrast. However, improving image quality, such as decreasing noise level and increasing contrast-to-noise ratio (CNR), without significantly prolonging scan time remains challenging.

Recently, deep learning (DL)-based reconstruction methods for MRI data have enabled image acceleration with substantially reduced image degradation in comparison with conventional parallel imaging and sparse-sampling image acceleration techniques, such as generalized autocalibrating partially parallel acquisitions (GRAPPA)<sup>9</sup> or compressed sensing.<sup>10</sup> These DL reconstruction techniques have been applied across a diverse array of sequences, anatomic regions, and MRI field strengths.<sup>11–14</sup> To date, there is limited validation of these techniques in clinical 7T imaging. This study aims to evaluate the performance of a DL image reconstruction technique, Deep Resolve Boost, applied to 2D T2-weighted TSE imaging at 7T. To our knowledge, this is the first study to clinically assess such *k*-space DL reconstruction techniques at 7T.

## MATERIALS AND METHODS

To assess improvements in image quality and presence of unique artifacts induced by DL reconstruction, we conducted a retrospective study of images acquired in a consecutive cohort of clinical patients who underwent 7T brain MRI for any indication between September 1, 2024 and October 20, 2024. Guidelines proposed by Strengthening the Reporting of Observational Studies in Epidemiology were followed (Supplemental Data). The raw *k*-space data were utilized to generate images through both DL-based reconstruction and standard “ground truth” reconstruction by using GRAPPA for comparative analysis. Patients were excluded from the analysis if pathologic changes were present in the areas of measurement that could potentially affect the

Received November 27, 2024; accepted after revision January 7, 2025.

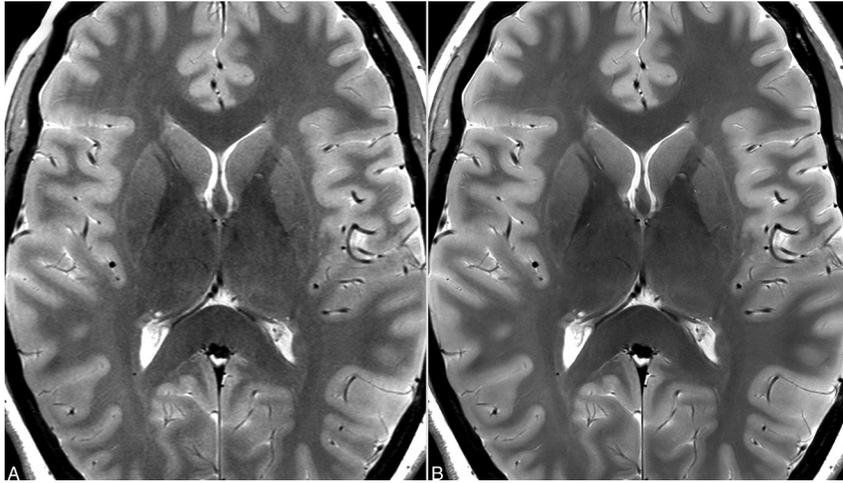
From the Department of Radiology (Z.L., X.Z., S.T., V.P., E.M.W., V.G., E.H.M.), Mayo Clinic, Jacksonville, Florida; Department of Radiology (Z.L.), Peking Union Medical College Hospital, Beijing, China; Siemens Medical Solutions USA, Inc (J.M.), Jacksonville, Florida; MR Application Predevelopment (D.N., P.L.), Siemens Healthineers AG, Forchheim, Germany; Siemens Healthineers (M.M.), Princeton, New Jersey; Siemens Medical Solutions USA, Inc (H.M.), Scottsdale, Arizona; and Department of Radiology (H.M.), Mayo Clinic, Scottsdale, Arizona.

Please address correspondence to Erik H. Middlebrooks, MD, Departments of Radiology and Neurosurgery, Mayo Clinic Florida, 4500 San Pablo Rd, Jacksonville, FL 32224; e-mail: middlebrooks.erik@mayo.edu; @EMiddlebrooksMD



Indicates article with supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8662>



**FIG 1.** A, Example axial 7T T2-weighted 2D TSE with standard reconstruction compared with the same data set reconstructed with deep learning (B).

expected normal structural homogeneity, such as demyelination of the corpus callosum.

### Image Acquisition

The 7T MRI scans were performed on a Magnetom Terra.X (Siemens Healthineers) by using an 8-channel transmit and 32-channel receive head coil (Nova Medical) operating in circularly polarized (“TrueForm”) B1 transmit mode. A 2D TSE T2-weighted sequence was acquired in the axial plane with key imaging parameters: TR = 5300 ms, TE = 30 ms, flip angle = 130°, echo-train length = 10, slice thickness = 2 mm, matrix = 576 × 370, pixel bandwidth = 280 Hz, FOV = 200 × 172 mm with an interpolated resolution of 0.17 × 0.17 mm, and 50 slices. A parallel imaging factor of 3 was applied with GRAPPA.

### DL Reconstruction

Deep Resolve Boost algorithm for 2D TSE imaging replaces traditional GRAPPA image reconstruction with a deep neural network trained on 7T data.<sup>11</sup> The network architecture resembles an iterative image reconstruction process, taking in undersampled raw data and pre-estimated coil sensitivity maps as input. After zero-filling, high-quality images are produced by alternating between a parallel imaging model and a DL-based regularization that enhances image quality. The key advantage of this technology is the reduction in acquisition time without sacrificing SNR or image quality, as supported by previous studies.<sup>11,13</sup> Image reconstruction results are subsequently enhanced by integrating Deep Resolve Boost with Deep Resolve Sharp (or super-resolution), which uses a deep neural network to increase the image resolution and sharpness.

For each sequence, the raw *k*-space data were reconstructed utilizing the proposed DL algorithms; subsequently, the raw data were retrospectively reconstructed by using the standard GRAPPA reconstruction pipeline to obtain “ground truth” images (Fig 1).

### Image Analysis

Quantitative and qualitative image analysis was performed to compare the DL images to the standard reconstructions. Qualitative

image analysis was independently performed by 2 subspecialty trained neuro-radiologists (E.H.M. and Z.L.) with 13 and 6 years of experience in radiology, respectively. For each data set, the standard and DL reconstructions were scored by using a 5-point Likert scale for: 1) overall image quality, 2) artifacts, 3) sharpness, 4) structural conspicuity, and 5) noise level.

Quantitative image analysis consisted of measures of CNR and image noise. An automated approach was used to standardize measures. First, the T2-weighted images were coregistered to a 3D T1-weighted magnetization-prepared 2 rapid acquisition gradient echo (MP2RAGE) uniform image, which was subsequently normalized to Montreal Neurological Institute (MNI) template space by using Advanced Normalization Tools (<https://stnava.github.io/ANTs/>). Next, 2 MNI coordinates were selected that included the center of the splenium of the corpus callosum and the head of the caudate. The MNI coordinate was then inversely warped into native patient space and a circular ROI with a 2.5 mm radius was expanded from each coordinate. Resulting ROIs were manually validated. Mean signal intensity and SD were calculated for each ROI. The SD of a homogeneous white matter area (splenium) was used as a noise estimate and CNR was calculated as:

$$CNR = \frac{|Splenium_{Mean} - Caudate_{Mean}|}{Splenium_{SD}}$$

### Statistical Analysis

Scores for image quality were not normally distributed and were presented by median with interquartile range in parentheses. The interobserver agreement between readers was assessed by  $\kappa$  analysis. The  $\kappa$  value was interpreted as follows: <0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; >0.80, almost perfect agreement. The Wilcoxon signed-rank test was used for comparison of image quality criteria between DL reconstruction and standard reconstruction.

A normal distribution of the quantitative data was confirmed by using the Shapiro-Wilk test. CNR and noise were compared between the DL and standard reconstruction by using a Wilcoxon matched-pairs signed rank test. A Bonferroni corrected *P* value of <.05 was considered statistically significant.

## RESULTS

Thirty-one consecutive subjects were identified. One subject was excluded due to severe demyelination of the splenium of the corpus callosum, yielding a total of 30 subjects included in the analysis. The mean age of the subjects was 55.8 years (range: 23–77), with 19 participants (63.3%) being women.

Image quality scores assessed by a 5-point Likert scale are shown in the Table. In both reader 1 and reader 2, DL-based

reconstruction has significantly higher image quality scores than standard reconstruction ( $P < .001$  for all).

The  $\kappa$  analysis for interobserver agreement demonstrated almost perfect agreement within DL-based reconstruction in structural conspicuity ( $\kappa = 0.813$ ), and substantial agreement in

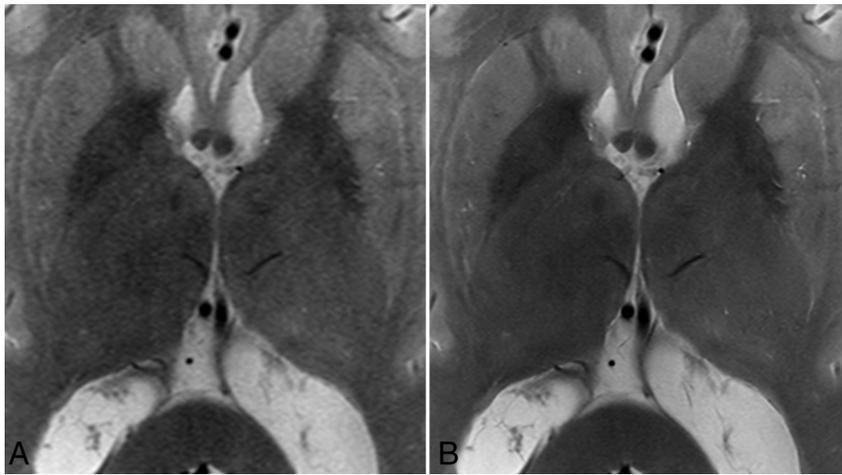
overall image quality ( $\kappa = 0.772$ ), artifacts ( $\kappa = 0.656$ ), sharpness ( $\kappa = 0.643$ ), and noise level ( $\kappa = 0.658$ ). The interobserver agreement within standard reconstruction reached substantial agreement in artifacts ( $\kappa = 0.714$ ), sharpness ( $\kappa = 0.669$ ), and noise level ( $\kappa = 0.628$ ), and moderate agreement in overall image quality ( $\kappa = 0.567$ ) and structural conspicuity ( $\kappa = 0.565$ ). Example images are shown in Fig 2.

The DL reconstruction resulted in a significant increase in CNR (Fig 3A) from a mean of 6.5 (95% CI 5.6–7.4) to 9.7 (95% CI 8.4–11.0) ( $P < .0001$ ). The DL reconstruction also resulted in a significant decrease in noise (Fig 3B) from a mean of 44.6 (95% CI 42.0–47.1) to 28.9 (95% CI 26.9–30.9) ( $P < .0001$ ). Using the DL reconstruction increased the CNR by a mean of 50.8% (95% CI 43.0–58.6%) and decreased the noise by a mean of 35.1% (95% CI 32.7–37.6%) (Fig 3C).

#### Image quality assessed by 5-point Likert scale

	Reader 1			Reader 2		
	DL	Standard	P Value	DL	Standard	P Value
Overall image quality	5 (4, 5)	4 (3.75, 4)	<.001	5 (4, 5)	4 (3, 4)	<.001
Artifacts	5 (4, 5)	4 (3, 4.25)	<.001	5 (4, 5)	4 (3, 4)	<.001
Sharpness	5 (4.75, 5)	4 (4, 4)	<.001	5 (4.75, 5)	4 (3, 4)	<.001
Structural conspicuity	5 (4.75, 5)	4 (4, 4)	<.001	5 (5, 5)	4 (4, 4)	<.001
Noise level	5 (4, 5)	4 (3, 5)	<.001	5 (4, 5)	4 (3, 4.25)	<.001

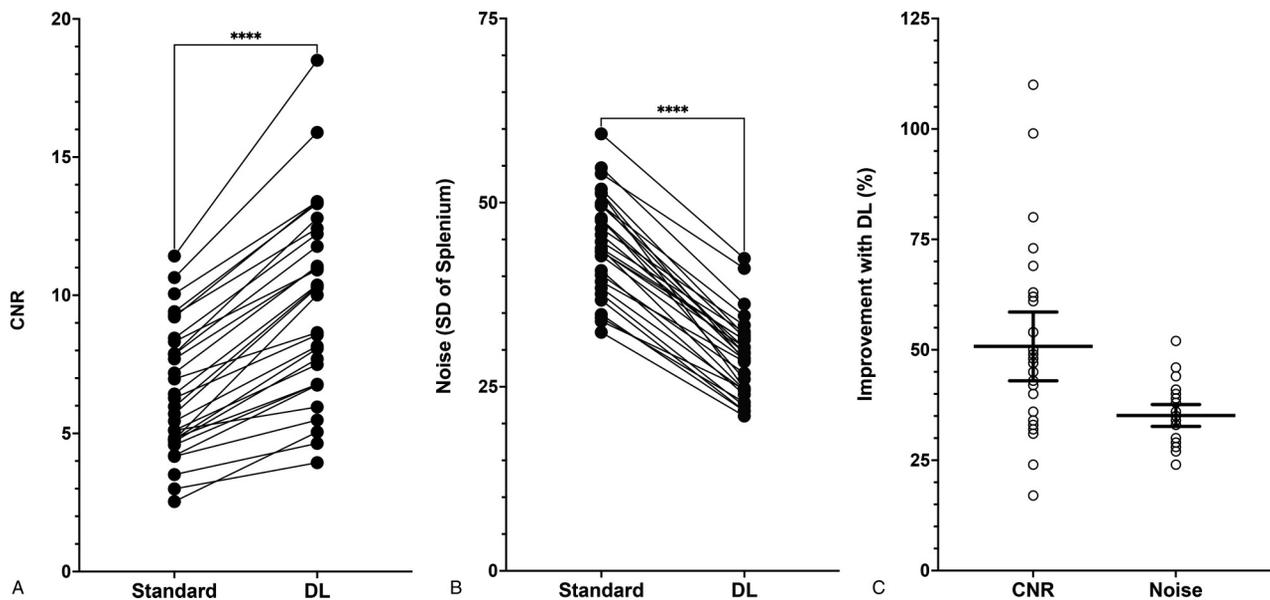
Data were presented as median with interquartile in parentheses.



**FIG 2.** A, Magnified view of an axial 7T T2-weighted 2D TSE with standard GRAPPA reconstruction. B, The same  $k$ -space data reconstructed with DL showing increased SNR, CNR, and sharpness.

#### DISCUSSION

This study provides validation of a DL-based reconstruction algorithm for 2D TSE at 7T. Compared with conventional GRAPPA reconstructions, the DL reconstruction resulted in a mean improvement in CNR of 50.8%. The DL images were also judged to have increased sharpness and improved structural conspicuity, all with no addition of image artifacts.



**FIG 3.** A, Paired line plot shows the change in CNR between the standard image reconstruction and the DL reconstruction. B, Paired line plot shows the change in image noise between the standard image reconstruction and the DL reconstruction. C, Percentage improvement in CNR and noise with the DL reconstruction (long bar = mean; whiskers = 95% CI).

These DL reconstructions have the potential to allow reductions in scan time, improved image quality, higher spatial resolution, and decreased specific absorption rate (SAR) in 7T imaging.

DL-based reconstructions have been implemented across nearly all anatomic regions, consistently demonstrating significant reductions in scan time, enhanced image quality, and diminished motion and other artifacts. Nevertheless, DL-based reconstructions encounter additional challenges at 7T. This study illustrates the advantages of DL reconstruction at 7T, which aims to reduce noise and enhance sharpness in reconstructed images by more intelligently modeling noise introduced by sparse  $k$ -space sampling. We observed substantial improvements in CNR and noise reduction compared with conventional acceleration techniques, without the introduction of artifacts or other detrimental effects.

DL-based reconstruction has the potential to facilitate higher spatial resolution by effectively mitigating the increased noise associated with smaller pixel sizes. Along the same lines, reducing SAR often necessitates modifications to pulse sequence parameters, such as an increase in echo-train length or repetition time, which consequently extends scan durations. To ensure clinically acceptable acquisition times, higher acceleration factors are generally employed; nonetheless, this approach introduces additional image noise. DL reconstruction may provide a viable solution by allowing the use of low-SAR sequence parameters and higher acceleration factors while maintaining an acceptable image SNR and acquisition time.

There are notable limitations to our study. First, to validate image effects strictly attributable to the DL reconstruction, we opted to reconstruct the raw  $k$ -space data from a standard clinical scan rather than perform repeat acquisitions, which could introduce effects not present in both scans (eg, patient motion). Consequently, we did not fully explore the potential range of the DL reconstruction, such as extending the limits of image time reduction or image resolution. Further tests will be necessary to determine the upper limits of image acceleration achievable through DL reconstruction. Second, although the algorithm should be agnostic to the contrast of the scan, we only directly tested T2-weighted images. Third, DL-based reconstructions generally require increased computational power that can increase reconstruction times. With the current protocol, slightly increased reconstruction times did not present a meaningful workflow change. Last, our sample size may not fully account for the range of imaging variations and artifacts that could be encountered in clinical scanning, necessitating future studies in larger sample sizes.

## CONCLUSIONS

DL reconstruction techniques in MRI can substantially improve scan time, resolution, and image quality, but have not been well

validated at 7T. We show that DL reconstruction of 2D TSE T2-weighted imaging at 7T can significantly enhance image quality without adverse effects.

Disclosure forms provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Park JE, Cheong EN, Jung DE, et al. **Utility of 7 Tesla magnetic resonance imaging in patients with epilepsy: a systematic review and meta-analysis.** *Front Neurol* 2021;12:621936 CrossRef Medline
2. Middlebrooks EH, Patel V, Zhou X, et al. **7 T lesion-attenuated magnetization-prepared gradient echo acquisition for detection of posterior fossa demyelinating lesions in multiple sclerosis.** *Invest Radiology* 2024;59:513–18 CrossRef Medline
3. Okromelidze L, Patel V, Singh RB, et al. **Central vein sign in multiple sclerosis: a comparison study of the diagnostic performance of 3T versus 7T MRI.** *AJNR Am J Neuroradiol* 2023;45:76–81 CrossRef Medline
4. Lakhani DA, Middlebrooks EH. **7-T neuromelanin and R2\* MRI in Parkinson disease.** *Radiology* 2022;305:296 CrossRef Medline
5. Lakhani DA, Zhou X, Tao S, et al. **Diagnostic utility of 7T neuromelanin imaging of the substantia nigra in Parkinson's disease.** *NPJ Parkinsons Dis* 2024;10:13 CrossRef Medline
6. Middlebrooks EH, Tipton PW, Greco E, et al. **Enhancing outcomes in deep brain stimulation: a comparative study of direct targeting using 7T versus 3T MRI.** *J Neurosurg* 2024;141:252–59 CrossRef Medline
7. Lakhani DA, Zhou X, Tao S, et al. **Clinical application of ultra-high resolution compressed sensing time-of-flight MR angiography at 7T to detect small vessel pathology.** *Neuroradiol J* 2023;36:335–40 CrossRef Medline
8. Cosottini M, Calzoni T, Lazzarotti GA, et al. **Time-of-flight MRA of intracranial vessels at 7 T.** *Eur Radiology Exp* 2024;8:68 CrossRef Medline
9. Griswold MA, Jakob PM, Heidemann RM, et al. **Generalized autocalibrating partially parallel acquisitions (GRAPPA).** *Magn Reson Med* 2002;47:1202–10 CrossRef Medline
10. Lustig M, Donoho D, Pauly JM. **Sparse MRI: the application of compressed sensing for rapid MR imaging.** *Magn Reson Med* 2007;58:1182–95 CrossRef Medline
11. Gassenmaier S, Afat S, Nickel D, et al. **Deep learning-accelerated T2-weighted imaging of the prostate: reduction of acquisition time and improvement of image quality.** *Eur J Radiol* 2021;137:109600 CrossRef Medline
12. Gassenmaier S, Afat S, Nickel MD, et al. **Accelerated T2-weighted TSE imaging of the prostate using deep learning image reconstruction: a prospective comparison with standard T2-weighted TSE imaging.** *Cancers (Basel)* 2021;13:3593 CrossRef Medline
13. Herrmann J, Koerzdoerfer G, Nickel D, et al. **Feasibility and implementation of a deep learning MR reconstruction for TSE sequences in musculoskeletal imaging.** *Diagnostics (Basel)* 2021;11:1484 CrossRef Medline
14. Radmanesh A, Muckley MJ, Murrell T, et al. **Exploring the acceleration limits of deep learning variational network-based two-dimensional brain MRI.** *Radiology Artif Intell* 2022;4:e210313 CrossRef Medline