Reader Instructions for Reporting Numeric Scores in Human Observer Studies for Unenhanced Head CT

The Discovery Workstation will ask you to record numeric confidence scores as you circumscribe imaging findings during image interpretation. This document provides important instructions on the meaning of the scores you will be asked to assign. These numeric scores play a critical role in the analysis of the study data, so care must be taken to ensure reproducibility of the ratings from one case to another and from one reading session to another.

Central Definitions

Primary Task. Each workstation instance is configured around a particular diagnostic task, which will be clearly identified within the system. For head CT in this study, the primary diagnostic task is identifying imaging findings associated with one of four causes of acute neurologic deficit (listed below) that may require further evaluation or treatment, or which may potentially explain patient signs and symptoms. The four target diagnoses under consideration in this study include:

- Infarction: acute, subacute, chronic, or indeterminate age
- Contusion
- Hemorrhage (nontraumatic intra-axial)
- Mass
- Extra-axial hemorrhage (subdural, subarachnoid, epidural, intraventricular).

Because neuroradiologists also need to consider normal aging processes and findings that may not correlate with patient symptoms, the following findings should not be marked and will be excluded for study purposes:

- Lacunar infarcts (even though it may potentially explain patient symptoms)
- Small-vessel ischemic change (leukoaraiosis)
- Benign intraparenchymal calcification
- Arachnoid cyst

For example, if the reader detects an acute brain infarction, the reader should mark regions in the brain that were suspicious for this infarction. Benign findings outside the brain and meninges would not require annotation, eg, calvarium, paranasal sinuses. Multiple findings may be present within 1 patient (eg, nontraumatic intra-axial hemorrhage and extra-axial hemorrhage; another example might be a dense MCA sign and the insular ribbon sign in acute infarction). Readers should circle each finding associated with these diagnoses (up to 5 findings). When you review images, the primary task should remain the focus of your examination. It is acknowledged that this narrow focus may provide a nonclinically relevant review, but it is essential to remain focused so that the technologies can be uniformly evaluated.

Lesion-Level Primary Task Confidence

For every lesion detected, you will be asked to give a diagnosis. You will then be asked to assign a confidence score as to the accuracy of this diagnosis using the same 0–100 scale. The important distinction here is that the primary task confidence rating reflects your confidence in your diagnosis attributed to this lesion and takes into account whether you think the lesion is present. For head CT, the lesion-level primary task confidence should reflect your confidence that one of the specific target findings (infarct, contusion, hemorrhage [intra-axial], mass, hemorrhage [extra-axial]) is present.

The lesion-level primary task confidence score is the most critical score that is assigned during the reading process. This numeric value will be used to summarize overall performance of the imaging configuration, so care must be taken to ensure proper calibration of the markings. This numeric rating will work in concert with the diagnosis code.

The lesion-level confidence score should reflect your confidence that one of the specific target findings (ie, infarct, contusion, intra- or extra-axial hemorrhage, or mass) is present within the circumscribed ROI. It should take into account whether the lesion is present. For example, if you are looking at noisy images and wonder if there is an infarct, you might assign a detection confidence of 50. If on the basis of the CT appearance, you believe the finding is most likely an infarct if there is an abnormality present, you might assign a lesion-level primary task confidence score of 40–45. In general, your lesion-level primary task confidence score (confidence that the selected target finding is present) should be less than or equal to your detection score (confidence that anything other than normal brain is present).

Your lesion-level primary task confidence score reflects your confidence that any of the target findings are present. For example, if you see a finding that is most likely a mass but could only be an infarct if not a mass, your confidence would be high because both of these possibilities are target findings.

- A confidence score of zero will be treated the same as if you did not mark the lesion, so if there is true suspicion of a diagnosis, the score should be one or greater. A score of zero would rarely be used, but we leave this as an option.
- If you have a high degree of confidence that 1 of the 5 target lesions is absent, consider assigning a numeric score from 1 to 25.
- If you have a high degree of confidence that the finding represents 1 of the 5 target lesions, consider assigning a confidence score from 75 to 100.
- One hundred is the highest lesion level primary task confidence score for a detection. It indicates that the circumscribed lesion is the most certain you have ever been that the lesion is present and that it represents the one of the primary diagnostic targets under study.

Final Considerations on Being Quantitative

You should attempt to make your confidence scores as quantitative as possible in terms of relative comparisons: A lesion given a particular confidence score of 45 should be more suspicious than a lesion given a confidence score of 35. For example, if you recall another lesion that you assigned a lesion-level confidence score of 80 while reviewing a new case that evokes a similar score in your mind but you believe this patient is a bit more

suspicious, your lesion-level confidence score should be slightly higher, say 83, to reflect this belief. Scoring is challenging and unfamiliar; just do your best.

If you use a confidence score rating more than once, you are indicating that there are no features indicating that 1 case (or ROI) is more suspicious than another.

You should try to space out your confidence score ratings to allow new cases and ROIs that have numeric scores between them.

Being quantitative is not easy. Do your best. If you are only comfortable using 10, 20, 30, and so forth, that is fine.