Evaluation of Lower Dose Spiral Head CT for Detection of Intracranial Findings Causing Neurologic Deficit

Joel G. Fletcher, M.D.[*,1], **David R. DeLone, M.D.**[1], **Amy L. Kotsenas, M.D.**[1], **Norbert G. Campeau, M.D.**[1], **Vance T. Lehman, M.D.**[1], **Lifeng Yu, Ph.D.**[1], **Shuai Leng, Ph.D**[1], **David R. Holmes III, Ph.D.**[2], **Phil Edwards**[2], **Matthew P. Johnson, M.S.**[3], **Gregory J. Michalak, Ph.D.**[1], **Rickey E. Carter, Ph.D.**[4], **Cynthia H. McCollough, Ph.D**[1]

[1]Department of Radiology

[2]Biomedical Imaging Resource

[3]Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN 5905

[4]Health Sciences Research, Mayo Clinic, Jacksonville, FL 32224

Abstract

Background and Purpose: To evaluate the performance of lower dose head CT for detection of intracranial findings resulting in neurologic deficit.

Materials and Methods: Projection data from 83 patients undergoing unenhanced spiral head CT for suspected neurologic deficit was collected. Positive cases required confirmation by histopathology, surgery, progression of findings, or corresponding neurological deficit; negative cases required negative neurological assessment. Routine dose was obtained using 250 effective mAs (emAs) and iterative reconstruction (IR). Lower dose configurations were reconstructed (25 emAs-IR, 50 emAs-filtered back projection (FBP) and -IR, 100 emAs-FBP and -IR, 200 emAs-FBP). Three neuroradiologists circled findings, indicating diagnosis, confidence (0 – 100), and image quality. The difference between the JAFROC figure-of-merit (FOM) at routine and lower dose configurations was estimated. A lower 95% CI estimate of the difference greater than −0.10 indicated non-inferiority.

Results: 42/83 patients had 70 intracranial findings (29 infarcts, 25 masses, 10 extra- and 6 intra-axial hemorrhages,) at routine head CT (CTDI 38.3 mGy). Routine dose JAFROC FOM was 0.87 (95% CI: 0.81, 0.93). Non-inferiority was shown for 100 emAs-IR (FOM difference −0.04 (95% CI: −0.08, 0.004) and 200 emAs-FBP (−0.02 (95% CI: −0.06, 0.02), but not for 100 emAs-FBP [−0.06 (95% CI: −0.10, −0.02)] or lower dose levels. Image quality was better at higher dose levels and with IR (p<0.05).

Conclusions: Observer performance for dose levels using 100 to 200 eff mAs was non-inferior to that observed at 250 eff mAs with IR, with IR preserving non-inferiority at a mean CTDI of 15.2 mGy.

*Corresponding author: 200 First Street SW, Rochester, MN 55905, Phone: 507-284-2511, Fax: 507-266-3661, fletcher.joel@mayo.edu.

Introduction

Unenhanced head CT is frequently requested in the emergency department or inpatient setting to examine patients with suspected neurologic deficit as well as those undergoing recent trauma. Surprisingly, the technique is not standardized, and the radiation dose varies substantially between institutions [1]. Radiologists strive to acquire CT exams at the lowest dose that will answer the diagnostic question in accordance with the ALARA principle [2] and the justification that the diagnostic benefit to the patient outweighs the small theoretical risk [3, 4] of associated radiation injury. Unfortunately, trade-offs between observer performance and radiation dose are not well delineated, even for most common CT tasks, and this dearth of information likely contributes to the differences in doses between institutions.

When using CT to answer very specific clinical questions, dramatic dose reduction for high contrast detection tasks utilizing unenhanced head CT can be undertaken (e.g., rule out craniosynostosis or shunt check) [5–7]. However, there is little scientific data or established consensus for what is the lowest acceptable dose for head CT for general evaluation of suspected neurologic deficit. CT detection of early acute cerebral infarction, subtle intracranial hemorrhage, or intracranial masses is a more challenging and demanding diagnostic task, as these pathologies result in only subtle low contrast differences in the involved structures. Iterative reconstruction can provide non-inferior performance for high contrast detection tasks at lower doses; however, recent data in phantoms and abdominal CT suggest that the improvement in observer performance compared to filtered back projection (FBP) may be limited [8, 9, 10, 11].

The majority of studies examining lower dose head CT with iterative reconstruction have relied upon subjective comparisons of image quality or contrast-to-noise ratios rather than observer (i.e., radiologist) performance. Practical considerations such as collecting cases with proven imaging findings, obtaining CT images at multiple doses, and correlating imaging findings between dose levels and reference standards make such research challenging.

We have recently studied a small number of patients (n=43) with suspected neurologic deficit [12, 13]. This preliminary study defined lower dose levels that were unable to preserve observer performance, but included a mix of both subtle and very obvious cases, which may have affected our ability to discriminate observer performance between varying dose levels. The current study builds on these initial results to compare observer performance between routine and lower dose levels for the detection of visually challenging intracranial findings causing acute neurologic deficit in a larger number of patients, and addressing the ability of iterative reconstruction to assist with dose reduction.

Methods

Patients and Cases

The primary diagnostic task for this study was to identify imaging findings that may require further treatment or evaluation, or which may potentially explain patient signs or symptoms, in patients with suspected acute neurologic deficit. After approval by our institutional review

board, we archived CT image and projection data of patients who underwent clinically-indicated spiral head CT examination for suspected acute neurologic deficit, and who provided consent to the use of medical records for research purposes. CT image and projection data was collected between August 20, 2013 and May 3, 2014. Archived CT images were then reviewed by a board-certified reference neuroradiologist (DRD, with 19 years of experience as an attending neuroradiologist) to ensure inclusion and exclusion criteria were met for this retrospective, case-control study.

All patients underwent CT for suspected acute neurologic deficit. Inclusion criteria were different for positive and negative cases, but required reference criteria to be met. Positive cases were additionally required to have sufficient clinical evidence to meet reference standard criteria for at least one of four target conditions causing acute neurologic deficit: (1) infarction (acute, subacute, chronic, or indeterminate age), (2) intra-axial hemorrhage (e.g., contusion), (3) mass, or (4) extra-axial hemorrhage (e.g., subdural, subarachnoid, epidural, intraventricular). Reference standard criteria for accepting positive cases into the study required confirmation of an imaging abnormality on the CT exam by the reference radiologist, in addition to (1) clinical physician confirmation of neurologic deficit at physical examination corresponding to abnormal imaging findings on the index CT exam, (2) progression or confirmation of imaging findings on another imaging exam (e.g., subsequent MRI, CT perfusion), or (3) confirmation of imaging findings at surgery. The reference neuroradiologist, unblinded to all clinical information and all subsequent imaging and surgical reports, then marked CT images for all CT findings relating to the target diagnosis that were present within the imaged volume that met inclusion criteria using a specially designed computer workstation, tightly circumscribing each CT imaging finding relating to each diagnosis and documenting correlative reference information using standard menus. A region of interest was also placed within white or grey matter (as appropriate) to reflect the background CT number in which the target lesion was located.

Negative cases were required to have sufficient evidence that a suspected acute neurologic deficit was not present. Reference standard criteria for accepting negative cases into the study included (1) both the clinical neuroradiologist interpreting the head CT at the time of imaging as well as the unblinded reference neuroradiologist indicating that no imaging findings associated with the four target conditions were present, and (2) lack of focal neurologic findings on physical examination by the clinical attending neurologist. Both positive and negative cases demonstrating small vessel ischemic change associated with aging (leukoaraiosis) were noted.

To determine the lowest radiation dose at which observer performance was non-inferior to routine dose, we constructed the study cohort to display visually challenging imaging findings that might affect a radiologist's ability to detect the target diagnoses at different radiation dose levels. Obvious CT findings such as a large intracranial hemorrhage can be detected at even the 10 – 20% dose level [12], and inclusion of such obvious cases does not help discriminate between the diagnostic differences between dose levels. The non-reader reference neuroradiologist visually evaluated 857 CT exams (620 negative and 237 positive) that met all reference and inclusion criteria, and graded the conspicuity of abnormal imaging findings associated with the target diagnoses along a 4 point scale: (1) minimally evident

(e.g., minimal obscuration of the lentiform nucleus and insular cortex in new infarct that might be easily missed), (2) subtle, (e.g. more definite acute infarct or small metastases on non-contrast spiral head CT), (3) distinct abnormality with well-defined borders (e.g. small chronic infarct, small intraparenchymal hematoma), or (4) obvious finding (e.g. chronic territorial infarct, large acute intraparenchymal hematoma, or diffuse subarachnoid hemorrhage). In this manner, cases with obvious imaging findings that would have no discriminatory value in selecting the appropriate radiation dose level would not be included. Based on this subjective conspicuity score, visually challenging positive cases that met reference criteria were selected from amongst the collected positive cases meeting inclusion criteria (Figure 1).

The target population for this study was constructed as previously described, with a population of 83 patients, approximately half of which would have one of the four target lesions causing acute neurologic deficit [12].

Image Acquisition and Reconstruction

Unenhanced spiral head CT exams were acquired using nearly identical CT platforms (Siemens Definition FLASH or Siemens Definition AS+; Siemens Healthineers) using a single x-ray tube, a detector configuration of 64×0.6 mm with a z-flying focal spot yielding 128 detector rows, 1 second tube rotation time, 120 kVp, and either 250 effective mAs (eff. mAs) or 340 eff. mAs. The higher tube current was obtained if the exam was part of a trauma scan in our emergency room. Routine dose CT images were reconstructed using a J40 head kernel using iterative reconstruction with a strength of 2 (SAFIRE; Siemens Healthineers, Malvern, PA), with 5 mm thick images reconstructed every 5 mm in the axial plane and 2 mm thick images reconstructed every 2 mm in the coronal plane. CT images corresponding to lower dose levels were created by inserting image noise in CT projection data using a highly accurate and validated noise insertion tool [5, 6, 13, 14]. For projection data obtained using 120 kV 340 eff mAs, noise was inserted to obtain images corresponding to 250 eff mAs, so that all patient "routine" dose exams corresponded to the same dose level. Based on the prior results, CT projection data dose levels corresponding to 200 eff. mAs, 100 eff. mAs, 50 eff. mAs, and 25 eff. mAs were then also created using iterative reconstruction at these dose levels, and additionally with FBP at 50 and 100 eff. mAs (Table 1). Noise-inserted CT projection data for each case was subsequently loaded back onto the CT system to reconstruct corresponding axial and coronal images with the appropriate reconstruction kernel.

Image Evaluation by Neuroradiologists

Three neuroradiologists with 18, 18 and 5 years of experience as clinical neuroradiologists at our institution were selected as blinded radiologist readers. Because of the unique features of head CT (complicated anatomy, variety of normal aging processes not representing pathology) a standardized reader training manual that defined pathologies to be detected and instructions for reporting reader confidence scores (with anchors) was developed and reviewed by each participating neuroradiologist (Appendix 1). Confidence scores ranged from 0 (indicating certainty the circumscribed finding is not one of the target lesions) to 100

(indicating the highest degree of certainty that can be achieved with CT for one of the target findings) [15, 16]. Neuroradiologists were instructed not to mark frequently seen aging processes such as small vessel ischemic change (leukoaraiosis) benign intraparenchymal calcification; chronic lacunar infarctions; or arachnoid cysts. Formal one-on-one reader training with the principal investigator was completed, with each reader interpreting 20 training cases selected to match the case mix, pathologies, and dose-reconstruction configurations in the subsequent reader study, discussing reader confidence ratings and any questions [17].

Readers evaluated routine dose-reconstruction and six lower dose-reconstruction configurations using a specialized computer workstation viewing images using at least two window settings (80/40 and 33/40), in multiple sessions. Once a CT finding corresponding to a target diagnosis was identified, readers were instructed to tightly circumscribe all imaging abnormalities corresponding to one the target diagnoses using a spline tool. Readers rated their level of confidence that one of the target diagnoses were present. Subsequently, readers answered four image quality questions related to overall quality, image sharpness, image noise and noise texture based on modified European Quality Criteria, with overall image quality rated along a 5-point scale (1 = nondiagnostic due to excessive noise or artifacts; 2 = diagnosis questionable due to excessive noise or artifacts, moderate decrease in confidence; 3 = diagnostic with moderate but acceptable noise or artifacts; 4 = mild noise, no change in confidence; and 5 = routine diagnostic image quality) [14, 18]. Exams were interpreted in random order, with only one dose-reconstruction configuration per patient interpreted during each session.

Statistical Analysis

The sample size for this study was determined as a part of a three-stage study design with the objective in this stage to screen and prioritize imaging strategies for evaluation in a large, future ten-reader, multicase study (stage 3). Stage 1, which consisted of 43 independent cases, has been previously published [12]. The original sample size calculations determined 83 cases were needed for this second stage of the study.

Matching of co-localized reference and reader detections was performed by the reference neuroradiologist (Figure 2). Comparison of reader performance between routine dose head CT (at 250 eff. mAs) using a head IR kernel (J40 strength 2) to lower dose-reconstruction configurations was performed using reader agreement rules and Jackknife alternative free-response receiver operating characteristic figure of merit (JAFROC FOM) non-inferiority analysis. Reader agreement rules compared reader localizations on the routine dose to the lower dose configurations. For positive cases, readers had to localize all target lesions, which were identified by two of three neuroradiologist readers at routine dose-configuration (denoted as "essential lesions"). For negative cases, no false positive localizations could be made by two or more of the neuroradiologist readers. Preset criteria for prioritization of a lower dose-configuration were agreement with routine dose interpretation in 71 of the 83 cases (86% of exams) [8].

JAFROC FOM analysis used the reader confidence scores input by the readers as well as their circumscribed imaging findings. A full description of JAFROC FOM for a mixed population as in this study has been provided previously [19, 20]. When more than one imaging finding was present in positive cases, findings were weighted according to the reciprocal of the number of findings. FOMs were calculated for every dose level and every reader. The contrasts (comparisons) of FOMs were estimated using the Hillis improvement [Dorfman, Berbaum and Metz (DBM) method under the modeling assumption of fixed readers - random cases using the Rjafroc package v1.0.1 on R version 3.4.2 (Vienna, Austria). Non-inferiority of lower dose configurations was represented by calculating the estimated difference between routine and lower dose configurations, with the limit of non-inferiority set at −0.10, a value determined a priori based on investigator consensus. This limit means that when the lower limit of the 95 percent confidence interval is greater than −0.10, non-inferiority with observer performance at routine dose is achieved.

In addition to these analyses, typical measures of diagnostic accuracy such as per-patient sensitivity and specificity and per-lesion sensitivity were performed. For the purposes of this calculation, a cut-off reader confidence of 10 (out of 100) for one of the target findings was used. For per-patient specificity, there could be no reader circumscriptions with a confidence level greater than 10 in negative cases. For these measures of diagnostic performance, generalized estimating equations (GEE) were used to estimate the pooled estimate across the three readers for each imaging strategy.

For the qualitative image quality ratings (e.g., overall impression of diagnostic image quality), a summary score was computed for each of the datasets. The summary score was the mean of the three readers. Tests for differences in image quality across dose and reconstruction were facilitated by a mixed model consisting of a random patient factor in order to account for the correlation of image qualities across the doses and assumed Gaussian errors. Post hoc comparisons of the quality summary score were considered descriptive and were not adjusted for multiple testing across doses. We evaluated for the effect of certain CT lesion characteristics such as size, CT number difference, and contrast-to-noise ratio compared to adjacent normal-appearing brain parenchyma on observer performance using Spearman rank coefficients comparing these parameters to mean reader confidence for correctly detected lesions. For false negative exams, a confidence score of 0 was assigned.

Results

Eighty-three CT exams in patients with suspected neurologic deficit hada mean $CTDI_{vol}$ of 38.1 ± 1.3 mGy. Forty-two positive cases had 70 proven target lesions (i.e., mean 1.6 ± 1.2 target imaging findings causing suspected neurologic deficit per patient). There were 29 infarcts, 25 masses, 10 extra-axial hemorrhages, and 6 intra-axial hemorrhages (Table 2). The mean area for the imaging findings was 6.9 ± 8.2 cm^2, with the longest and shortest linear dimensions being 3.9 ± 3.0 cm and 2.2 ± 1.1 cm, respectively. Thirty patients (36%) had leukoaraosis (17 with proven target lesions, 13 without lesions).

Table 3 shows reader agreement for lower dose configurations along with JAFROC FOM's. The 25 eff. mAs IR configuration failed to meet preset criteria for reader agreement rules. For lower dose configurations, a greater proportion of non-agreement with routine dose comes from missed lesions (false negatives interpretations) rather than false positive interpretations (or localizations) in negative exams (Figures 2 and 3). At least 2 of 3 of the neuroradiologists identified 94% of the target lesions at 200 eff. mAS/FBP, and this declined to 87% at 50 eff. mAS IR and 81% at 50 eff. FBP while the number of correctly interpreted negative cases remained virtually identical.

Table 4 shows the generalized estimating equations (GEE) per-patient sensitivity and specificity, as well as the target lesion sensitivity for CT finding accounting for neurologic deficit in our patient population. Per-target lesion GEE sensitivity was significantly decreased compared to routine dose unenhanced head CT for 25 eff. mAs IR configuration (Figures 4 and 5). Per-patient GEE sensitivity was not significantly different from the reference dose at any lower configurations. GEE per patient specificity declined only slightly at lower doses.

Figure 4 is the forest plot demonstrating the difference between routine dose and the lower dose configurations. Only the 100 eff. mAs IR and 200 eff. mAs FBP lower dose configurations demonstrated non-inferiority compared to routine dose, with non-inferiority not demonstrated for 100 eff. mAS FBP and lower dose configurations.

Figure 5 demonstrates the overall image quality ratings for routine and lower dose configurations. Each lower dose configuration had significantly inferior image quality ($p < 0.05$). At 100 eff mAs and 50 eff mAs dose levels, configurations with iterative reconstruction had significantly higher image quality ($p < 0.0001$).

Mean reader confidence was moderately affected by lesion size (Spearman correlation coefficients for lower dose configurations: $p = 0.23$, $p = 0.053$ for FBP 200 eff. mAs and $p = 0.35 - 0.39$, $p < 0.05$ for lower doses). Mean reader confidence had a somewhat weaker relationship with both contrast to noise ratio ($p = 0.16$, $p = 0.18$ for FBP 200 eff. mAs and $p = 0.24 - 0.32$, $p < 0.05$ for lower doses) and CT number difference ($p = 0.11$, $p = 0.35$ at FBP 200 eff. mAs, $p = 0.2 - 0.29$ for lower doses with $p < 0.05$ at FBP 100 eff. mAs, FBP 50 eff. mAs and IR 25 eff. mAs).

Discussion

In keeping with the ALARA principle, radiologists strive to perform diagnostically useful imaging at the lowest possible radiation exposure dose to the patient. In this work, we systematically evaluated the ability to reduce radiation dose for spiral unenhanced head CT for acute neurologic deficit without compromising neuroradiologist observer performance. Our study cohort included patients with suspected neurologic deficit with either proven positive CT findings by reference criteria or negative CT exams accompanied by lack of focal neurologic findings on physical examination by the clinical attending neurologist. We found that there were small, non-significant decreases in sensitivity for CT findings accounting for neurologic deficit at all lower dose configurations, but that there was

substantial opportunity for dose reduction if small decreases in performance might be acceptable. Specifically, we found observer performance was non-inferior down to 40% of our routine dose level (i.e, a tube current setting of 100 effective mAs, corresponding to a $CTDI_{vol}$ of 15.2 mGy; Table 4) when iterative reconstruction was employed. At the same dose level with images reconstructed with weighted FBP, sensitivity for target lesions accounting for neurologic deficit declined by about 2–3% and specificity declined by 1.5–5%

Prior work examining the ability to lower radiation dose at unenhanced head CT has largely been focused on the ability to improve image quality [21–23], as this may be a barrier to radiologists examining noisier, lower dose images. Wu et al. examined image quality and diagnostic accuracy using iterative reconstruction using a case-control design and lowering rotation dose by 43% with either tube current or tube voltage reduction [24]. They found that dose reduction achieved with tube current reduction and IR preserved image quality, but their study did not report observer performance for pathologic entities, relying instead detection of subcortical arteriosclerotic encephalopathy and number of infarcts as a surrogate marker. Additionally, several studies evaluating unenhanced head CT have not used spiral head CT, which can facilitate lower radiation dose compared to sequential scanning [25]. Without observer performance data to ensure that lower doses do not result in inferior performance, routine diagnostic levels are often set by subjective image quality perceptions of individuals or by benchmarking to other institutions. Having observer performance data on which to facilitate dose reduction is needed for making objective decisions and may assist in overcoming differences between institutions.

Recent clinical and phantom studies in abdominal CT have highlighted the limitations of iterative reconstruction and its ability to facilitate radiation dose reduction for low contrast objects such as hepatic metastases. Mileto et al recently had 16 radiologists examine a low contrast detectability phantom and found that radiation doses resulted in similar declines in observer performance for FBP and IR, and that differences between radiologists were larger than across reconstruction algorithms [9]. Jensen et al. found that approximately 50% dose reduction with one iterative reconstruction approach did not preserve observer performance for the detection of hepatic metastases [11]. Fletcher et al. found that FBP and IR performed similarly at lower doses in detecting hepatic metastases, and that IR might only be needed to maintain observer performance at a certain threshold dose [8, 12]. Similar to these studies, we found that performance for the detection of intracranial findings causing acute neurologic deficit declined slightly at lower doses using JAFROC FOM as a metric of observer performance, but that iterative reconstruction was helpful in preserving non-inferiority at 100 eff. mAs. Based on this work, we plan to refine the results with a with a multi-reader, multi-case study with ten neuroradiologists in order to better predict the lowest dose that can preserve performance.

Our study has limitations. Because of the concern for missing subtle findings in patients with acute neurologic deficit, we used a retrospective case-control study design using an enriched cohort of visually challenging CT findings to discriminate between radiation dose levels, with these findings proven based on surgical assessment, follow-up imaging, or corresponding neurologic deficit. From a radiation exposure standpoint it is not possible to

reimage patients directly at multiple differing exposures during the same imaging session. Therefore, we relied on a validated noise insertion method to reconstruct CT images corresponding to multiple dose levels; however, we have found that this method is highly accurate, and has allowed our clinical practice to readily adopt research findings arrived at using this method [5, 6]. CT interpretation at reduced levels is subjectively more effortful and fatiguing than at standard doses, and the potential effects of radiologist fatigue could not be measured in our study as neuoradiologists interpreted cases with different dose levels in each interpretation session. Finally, our results relied upon interpretations by only three neuroradiologists, and extrapolation of their interpretations to a larger number of neuroradiologists or general radiologists may be limited.

Conclusion

This study helps to better define the potential dose reduction that can be achieved with conventional spiral CT and IR that will maintain diagnostic performance for evaluation of suspected neurologic deficits. Our study demonstrates that substantial opportunity exists for dose reduction using for spiral non-enhanced head CT, and that, dose level might potentially be reduced to 40% of routine dose levels or a CTDI$_{vol}$ of approximately 15 mGy if slight decreases in performance are acceptable (e.g., in follow up and surveillance). Furthermore, the beneficial effect of IR was most pronounced at this 15 mGy dose level. Above this dose level, observer performance might be preserved with filtered back projection alone.

Acknowledgments:

Appendix 1.: Reader instructions for reporting numeric scores in human observer studies for unenhanced head CT

The Discovery Workstation will assess up numeric confidence scores depending on the case and task. 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 specific findings (listed below) that may require further evaluation**

or treatment, or which may potentially explain patient signs and symptoms. Findings to be noted include:

- Infarction – acute, subacute, chronic or indeterminate age

- Contusion

- Hemorrhage (non-traumatic intra-axial)

- Mass

- Extra-axial hemorrhage (subdural, subarachnoid, epidural, intraventricular)

Because neuroradiologists also need to consider normal aging processes and findings which 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 in outside of the brain and meninges would not require annotation, e.g. calvarium, paranasal sinuses. Multiple findings may be present within one patient (e.g., non-traumatic intra-axial hemorrhage and extra-axial hemorrhage; another example might be a dense MCA sign and insular ribbon sign in acute infarction). Readers should circle each finding associated with these diagnoses (up to five 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 non-clinically 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 in 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 or not 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.

Lesion-level Primary Task Confidence Score

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 (i.e., infarct, contusion, intra- or extra-axial hemorrhage, or mass) is present within the circumscribed ROI. It should take into account whether or not 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 based on 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 mostly 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 0 will be treated the same as if you did not mark the lesion, so if there is true suspicion for a diagnosis, the score should be one or greater. A score of 0 would rarely be used, but we leave this as an option.

- If you have a high degree of confidence that one of the five target lesions is absent, consider assigning a numeric score from 1 – 25.

- If you have a high degree of confidence that the finding represents one of the five target lesions, consider assigning a confidence score from 75–100.

- 100 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 one case (or ROI) is more suspicious than another.

You should try and space out your *Confidence Score* ratings to allow for new cases and ROIs that have *Numeric Score* in between.

Being quantitative is not easy. Do your best. If you are only comfortable using 10, 20, 30, etc… That's fine.

Abbreviations

ALARA, CT	as low as reasonably acheivable
JAFROC FOM	Jackknife alternative free-response receiver operating characteristic figure of merit
IR	iterative reconstruction
FBP	filtered back projection
GEE	generalized estimating equations
emAs	effective mAs

References

1. Smith-Bindman R, Wang Y, Yellen-Nelson TR, et al. Predictors of CT Radiation Dose and Their Effect on Patient Care: A Comprehensive Analysis Using Automated Data. Radiology 2017;282:182–193 [PubMed: 27438166]

2. Centers for Disease Control and Prevention. ALARA - As Low As Reasonably Achievable. https://www.cdc.gov/nceh/radiation/alara.html. 2 22, 2019.

3. McCollough CH, Bushberg JT, Fletcher JG, et al. Answers to Common Questions About the Use and Safety of CT Scans. Mayo Clin Proc 2015;90:1380–1392 [PubMed: 26434964]

4. McCollough CH, Guimaraes L, Fletcher JG. In defense of body CT. AJR Am J Roentgenol 2009;193:28–39 [PubMed: 19542392]

5. Gabriel S, Eckel LJ, DeLone DR, et al. Pilot study of radiation dose reduction for pediatric head CT in evaluation of ventricular size. AJNR Am J Neuroradiol 2014;35:2237–2242 [PubMed: 25082822]

6. Montoya JC, Eckel LJ, DeLone DR, et al. Low-Dose CT for Craniosynostosis: Preserving Diagnostic Benefit with Substantial Radiation Dose Reduction. AJNR Am J Neuroradiol 2017;38:672–677 [PubMed: 28183836]

7. Widmann G, Schullian P, Gassner EM, et al. Ultralow-dose CT of the craniofacial bone for navigated surgery using adaptive statistical iterative reconstruction and model-based iterative reconstruction: 2D and 3D image quality. AJR Am J Roentgenol 2015;204:563–569 [PubMed: 25714286]

8. Fletcher JG, Fidler JL, Venkatesh SK, et al. Observer Performance with Varying Radiation Dose and Reconstruction Methods for Detection of Hepatic Metastases. Radiology 2018;289:455–464 [PubMed: 30204077]

9. Mileto A, Zamora DA, Alessio AM, et al. CT Detectability of Small Low-Contrast Hypoattenuating Focal Lesions: Iterative Reconstructions versus Filtered Back Projection. Radiology 2018;289:443–454 [PubMed: 30015591]

10. Goenka AH, Herts BR, Obuchowski NA, et al. Effect of reduced radiation exposure and iterative reconstruction on detection of low-contrast low-attenuation lesions in an anthropomorphic liver phantom: an 18-reader study. Radiology 2014;272:154–163 [PubMed: 24620913]

11. Jensen CT, Wagner-Bartak NA, Vu LN, et al. Detection of Colorectal Hepatic Metastases Is Superior at Standard Radiation Dose CT versus Reduced Dose CT. Radiology 2018:181657

12. Fletcher JG, Yu L, Fidler JL, et al. Estimation of Observer Performance for Reduced Radiation Dose Levels in CT: Eliminating Reduced Dose Levels That Are Too Low Is the First Step. Acad Radiol 2017;24:876–890 [PubMed: 28262519]

13. Yu L, Shiung MM, Jondal D, et al. Development and Validation of a Practical Lower-dose-simulation Tool for Optimizing CT Scan Protocols. J Comput Assist Tomogr 2012;36:477–487 [PubMed: 22805680]

14. Fletcher JG, Yu L, Li Z, et al. Observer Performance in the Detection and Classification of Malignant Hepatic Nodules and Masses with CT Image-Space Denoising and Iterative Reconstruction. Radiology 2015;276:465–478 [PubMed: 26020436]

15. Briggs M, Closs JS. A descriptive study of the use of visual analogue scales and verbal rating scales for the assessment of postoperative pain in orthopedic patients. J Pain Symptom Manage 1999;18:438–446 [PubMed: 10641470]

16. Hadjiiski L, Chan HP, Sahiner B, et al. Quasi-continuous and discrete confidence rating scales for observer performance studies: Effects on ROC analysis. Acad Radiol 2007;14:38–48 [PubMed: 17178364]

17. Dendumrongsup T, Plumb AA, Halligan S, et al. Multi-reader multi-case studies using the area under the receiver operator characteristic curve as a measure of diagnostic accuracy: systematic review with a focus on quality of data reporting. PLoS One 2014;9:e116018 [PubMed: 25541977]

18. Froemming AT, Kawashima A, Takahashi N, et al. Individualized kV selection and tube current reduction in excretory phase computed tomography urography: potential for radiation dose reduction and the contribution of iterative reconstruction to image quality. J Comput Assist Tomogr 2013;37:551–559 [PubMed: 23863531]

19. Chakraborty D, Yoon H. JAFROC analysis revisited: figure-of-merit considerations for human observer studies. Proc SPIE Med Imag 2009;7263:72630T

20. Chakraborty D, Zhai X. RJafroc: Modeling, Analysis, Validation and Visualization of Observer Performance Studies in Diagnostic Radiology. https://CRAN.R-project.org/package=RJafroc. 1 14, 2019.

21. Harris M, Huckle J, Anthony D, et al. The Acceptability of Iterative Reconstruction Algorithms in Head CT: An Assessment of Sinogram Affirmed Iterative Reconstruction (SAFIRE) vs. Filtered Back Projection (FBP) Using Phantoms. J Med IMag Rad Sci 2017;48:259–269

22. Korn A, Fenchel M, Bender B, et al. Iterative reconstruction in head CT: image quality of routine and low-dose protocols in comparison with standard filtered back-projection. AJNR Am J Neuroradiol 2012;33:218–224 [PubMed: 22033719]

23. Love A, Olsson ML, Siemund R, et al. Six iterative reconstruction algorithms in brain CT: a phantom study on image quality at different radiation dose levels. Br J Radiol 2013;86:20130388 [PubMed: 24049128]

24. Wu TH, Hung SC, Sun JY, et al. How far can the radiation dose be lowered in head CT with iterative reconstruction? Analysis of imaging quality and diagnostic accuracy. Eur Radiol 2013;23:2612–2621 [PubMed: 23645331]

25. Pace I, Zarb F. A comparison of sequential and spiral scanning techniques in brain CT. Radiol Technol 2015;86:373–378 [PubMed: 25835403]
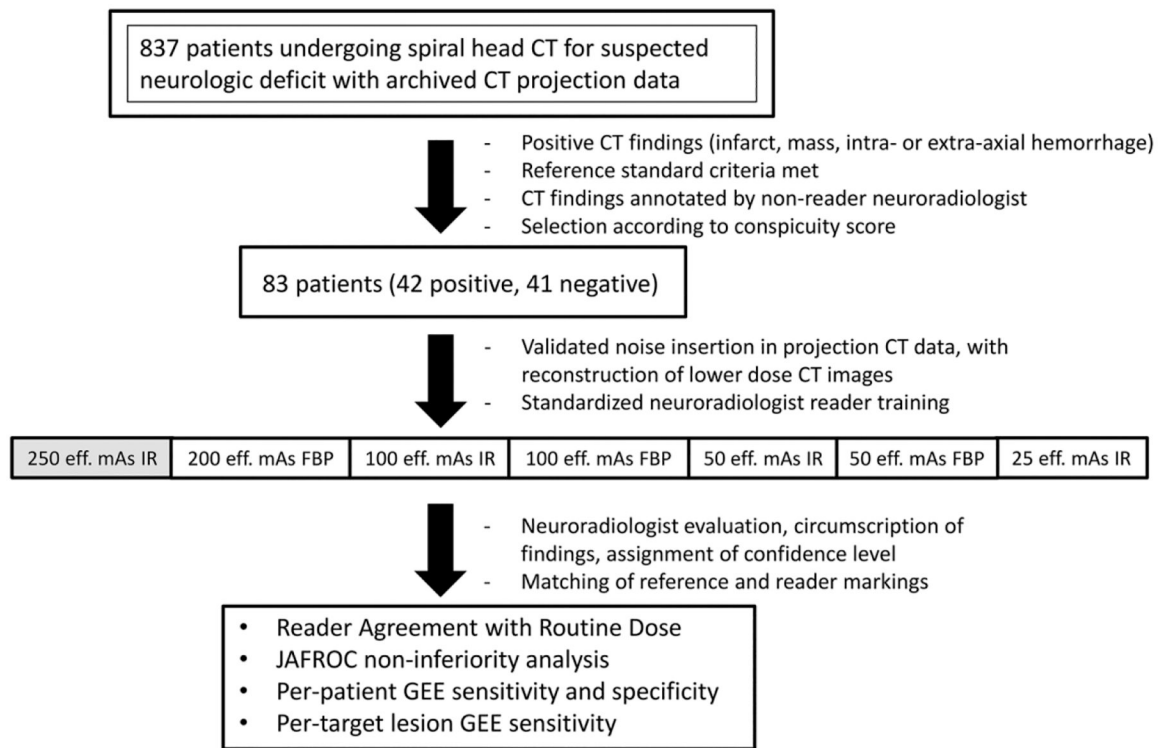
Figure 1.
Study schema. (eff. mAs = effective tube current-time product; FBP = filtered back projection; IR = iterative reconstruction; JAFROC = Jackknife alternative free-response receiver operating characteristic figure of merit; GEE = generalized estimating equations)
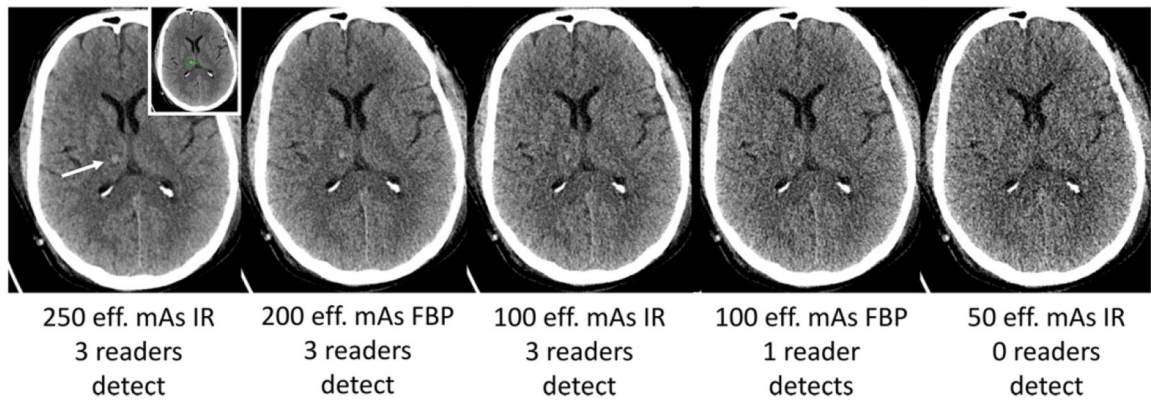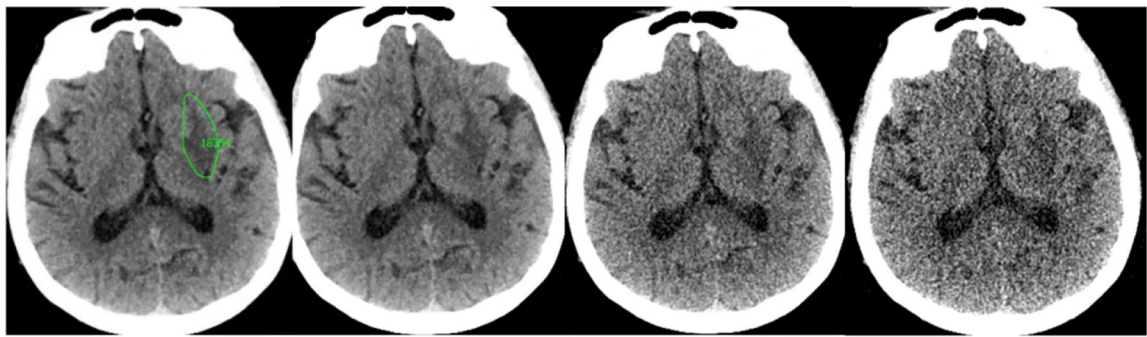
250 eff. mAs IR
3 readers
detect

200 eff. mAs FBP
3 readers
detect

100 eff. mAs IR
3 readers
detect

100 eff. mAs FBP
1 reader
detects

50 eff. mAs IR
0 readers
detect

Figure 2.
Small right thalamic hemorrhage (white arrow) shown on routine dose CT image (250 eff. mAs IR) along with lower dose configurations. Small left inset shows reference neuroradiologist marking of the target lesion (green circle). This CT exam was performed after trauma, with hemorrhage confirmed surgically, and the final diagnosis recorded as right thalamic hemorrhage consistent with shear injury.

250 eff. mAs IR
Routine Dose
3 readers
detect

200 eff. mAs FBP
3 readers
detect

100 eff. mAs FBP
3 readers
detect

50 eff. mAs FBP
2 readers

Figure 3.
Acute left lentiform nucleus infarct (green circle indicates reference neuroradiologist marking at routine dose) with corresponding images lower dose FBP CT images along with reader results. The imaging finding on this CT exam evolved over time with corresponding clinical confirmation of corresponding neurologic deficit by a staff neurologist, and the final diagnosis was recorded as acute left striatal infarct.
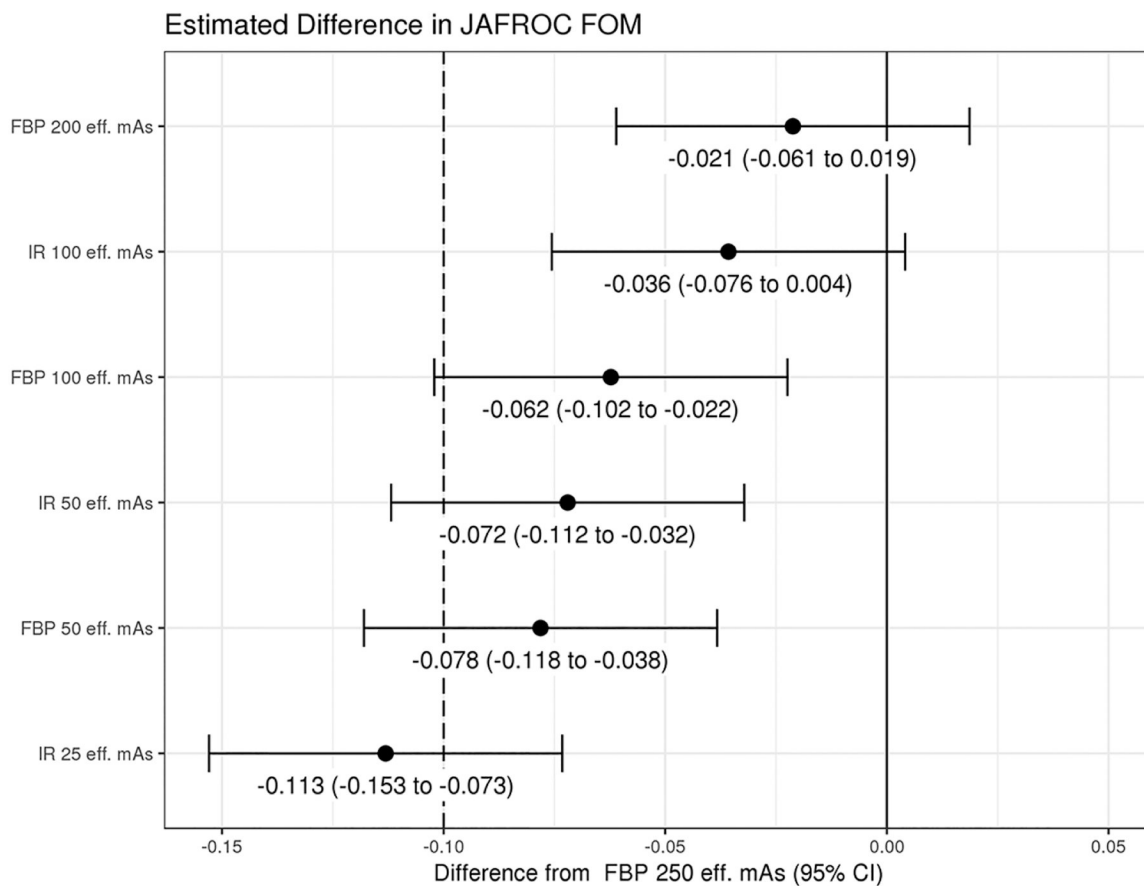
Figure 4.
Non-inferiority analysis showing the difference between JAFROC FOM at routine dose and the lower dose configurations for CT findings causing acute neurologic deficit. Limit of non-inferiority was established a priori to be −0.10, meaning that if the lower limit of the 95% confidence interval is greater than −0.10, then non-inferiority was shown. (FOM = figure of merit; JAFROC = Jackknife alternative free-response receiver operating characteristic)
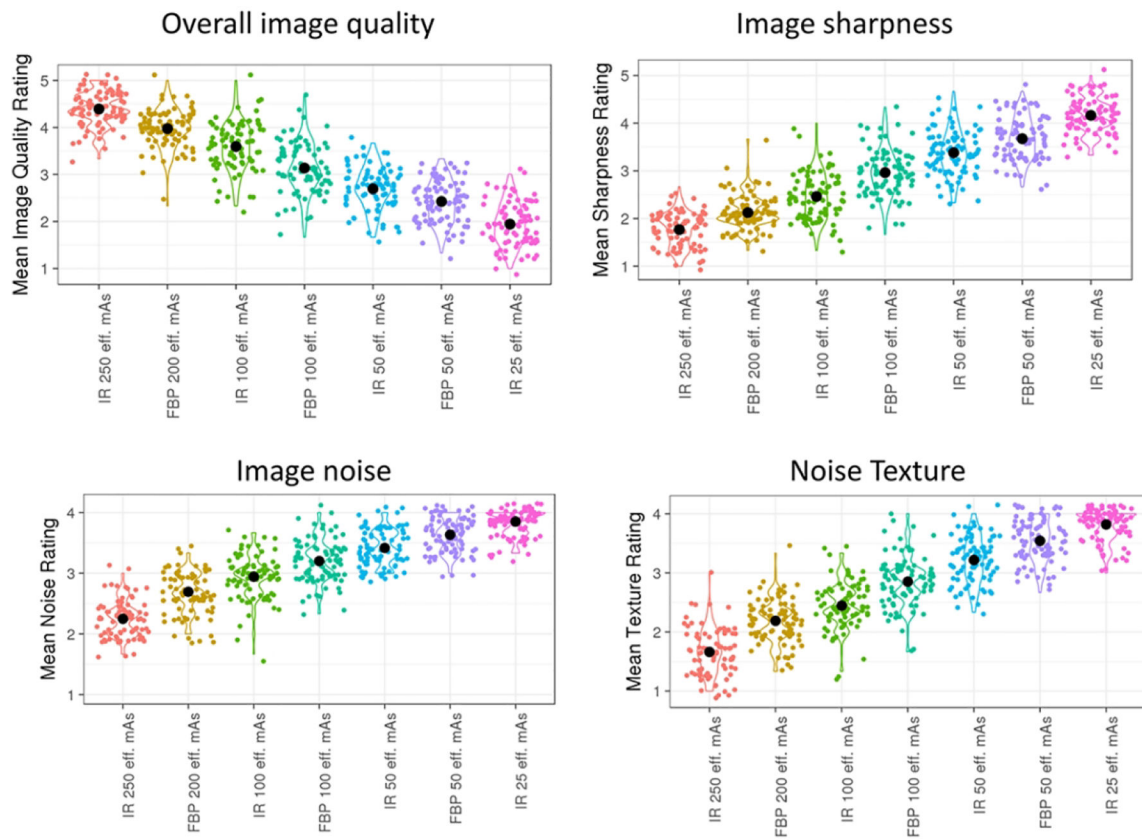
Figure 5.
Image quality metrics for routine and lower dose configurations in this study. Optimal ratings were 5 for image quality and 1 for individual image metrics.

Table 1.

Dose levels and reconstruction kernels for unenhanced CT examinations used in this study.

Tube Current Setting (effective mAs)	$CTDI_{vol}$ (mGy)	Reconstruction Kernel (Type[*], Strength)
250 eff. mAs	38.1	J40 (IR – 2)
200 eff. mAs	30.5	J40 (IR – 2)
100 eff. mAs	15.2	J40 (IR – 2)
100 eff. mAs	15.2	H40 (FBP)
50 eff. mAs	7.6	J40 (IR – 2)
50 eff. mAs	7.6	H40 (FBP)
25 eff. mAs	3.8	J40 (IR – 2)

[*] FBP = filtered back projection; IR = iterative reconstruction; eff. mAs = effective mAs

Table 2.

Reference documentation and conspicuity of proven lesions in positive CT exams with imaging findings corresponding to cause of acute neurologic deficit (n=42).

Target Diagnosis	Number of Imaging Findings with Target Diagnosis	# with reference criterion (non-exclusive list)	Ranking of conspicuity scores[a] Mean (SD)
Infarct	29	• Clinical confirmation of corresponding deficit = 29 • Progression/confirmation on another imaging study= 23 • Confirmation at surgery= 0	2.10 (0.76)
Mass	25	• Clinical confirmation of corresponding deficit = 22 • Progression/confirmation on another imaging study= 21 • Confirmation at surgery= 2	2.80 (0.70)
Extra-axial hemorrhage	10	• Clinical confirmation of corresponding deficit = 10 • Progression/confirmation on another imaging study= 3 • Confirmation at surgery= 0	2.60 (0.66)
Intra-axial hemorrhage	6	• Clinical confirmation of corresponding deficit = 6 • Progression/confirmation on another imaging study= 2 • Confirmation at surgery= 0	2.67 (0.47)

[a]Please see Methods; in brief, conspicuity scores: (1) minimally evident, (2) subtle, (3) distinct focal abnormality, (4) obvious.

Table 3.

Reader agreement of lower dose-reconstruction configurations compared to routine dose unenhanced head CT exams, along with JAFROC FOM's. The JAFROC FOM for routine unenhanced head CT (250 eff. mAs with IR) was 0.867 (0.805 – 0.929). (eff. mAs = effective milliampere-seconds, the tube current; FBP = filtered back projection; IR = iterative reconstruction)

Lower dose-reconstruction configuration	% of the 47 Essential Lesions[a] Detected by Readers at Lower dose configuration		# of Successful interpretations per lower dose Reconstruction-configuration			JAFROC FOM FOM (95% CI)
	2 of 3	3 of 3	Cases with at least one essential lesion n=34	Cases without any essential lesions n=49	Number Success ful Interpret ation (71 or more required per design)	
200 eff. mAs FBP	44 (94%)	39 (83%)	30	48	78	0.846 (0.780 – 0.912)
100 eff. mAs IR	43 (92%)	37 (79%)	29	46	75	0.831 (0.764 – 0.898)
100 eff. mAs FBP	42 (89%)	36 (77%)	28	45	73	0.805 (0.732 – 0.878)
50 eff. mAs IR	41 (87%)	32 (68%)	26	47	73	0.795 (0.727 – 0.864)
50 eff. mAs FBP	38 (81%)	31 (66%)	25	47	72	0.789 (0.717 – 0.861)
25 eff. mAs IR	34 (72%)	25 (53%)	22	45	67[b]	0.754 (0.681 – 0.827)

[a]Essential lesions are described in the Methods. Briefly, they represent lesions correctly localized and classified at the routine dose (250 eff. mAs with IR) by a majority of readers.

[b]Dose-reconstruction configuration did not meet preset criteria for agreement with routine dose interpretation, which was defined as agreement in 71 of the 83 exams.

Table 4.

Per patient and per lesion sensitivity and specificity using generalized estimating equations (GEE) for target neurologic findings account for acute neurologic deficit.

Dose-Kernel Configuration	Per-patient Sensitivity for CT Findings Accounting for Acute Neurologic Deficit GEE % (95% C.I.) (range%)	Per-patient Specificity for CT Findings Accounting for Acute Neurologic Deficit GEE % (95% C.I.) (range%)	Target Lesion Sensitivity for CT Findings Accounting for Acute Neurologic Deficit GEE% (95% C.I.) (range%)
250 eff. mAs IR	81.7 (71.1 to 92.3) (78.6, 83.3)	93.5 (88.9 to 98.1) (85.4, 100.0)	71.0 (64.8 to 77.1) (65.7, 74.3)
200 eff. mAs FBP	79.4 (68.2 to 90.6) (76.2, 83.3)	91.9 (87.5 to 96.3) (80.5, 100.0)	68.6 (62.3 to 74.9) (61.4, 72.9)
100 eff. mAs IR	77.0 (65.5 to 88.5) (73.8, 81.0)	88.6 (82.8 to 94.4) (73.2, 95.1)	68.1 (61.8 to 74.4) (64.3, 71.4)
100 eff. mAs FBP	74.6 (62.4 to 86.8) (69.0, 78.6)	87.0 (81.1 to 92.9) (75.6, 95.1)	62.9 (56.3 to 69.4) (57.1, 65.7)
50 eff. mAs IR	73.8 (62.3 to 85.4) (66.7, 78.6)	88.6 (82.8 to 94.4) (75.6, 97.6)	60.5 (53.9 to 67.1) (52.9, 65.7)
50 eff. mAs FBP	72.2 (60.6 to 83.8) (71.4, 73.8)	83.7 (77.7 to 89.8) (69.8, 86.0)	61.0 (54.4 to 67.6) (60.0, 62.9)
25 eff. mAs IR	65.9 (53.3 to 78.4) (61.9, 71.4)	88.6 (82.4 to 94.8) (85.4, 92.7)	53.3 (46.6 to 60.1) (45.7, 62.9) [a]

[a] 95% confidence interval does not overlap with routine dose, so dose-reconstruction configuration is significantly worse.