

# Radiology Resident Evaluation of Head CT Scan Orders in the Emergency Department

William K. Erly, William G. Berger, Elizabeth Krupinski, Joachim F. Seeger, and John A. Guisto

**BACKGROUND AND PURPOSE:** Prior studies have revealed little difference in residents' abilities to interpret cranial CT scans. The purpose of this study was to assess the performance of radiology residents at different levels of training in the interpretation of emergency head CT images.

**METHODS:** Radiology residents prospectively interpreted 1324 consecutive head CT scans ordered in the emergency department at the University of Arizona Health Science Center. The residents completed a preliminary interpretation form that included their interpretation and confidence in that interpretation. One of five neuroradiologists with a Certificate of Added Qualification subsequently interpreted the images and classified their assessment of the residents' interpretations as follows: "agree," "disagree-insignificant," or "disagree-significant." The data were analyzed by using analysis-of-variance or  $\chi^2$  methods.

**RESULTS:** Overall, the agreement rate was 91%; the insignificant disagreement rate, 7%; and the significant disagreement rate, 2%. The level of training had a significant ( $P = .032$ ) effect on the rate of agreement; upper-level residents had higher rates of agreement than those of more junior residents. There were 62 false-negative findings. The most commonly missed findings were fractures ( $n = 18$ ) and chronic ischemic foci ( $n = 12$ ). The most common false-positive interpretations involved 10 suspected intracranial hemorrhages and suspected fractures.

**CONCLUSION:** The level of resident training has a significant effect on the rate of disagreement between the preliminary interpretations of emergency cranial CT scans by residents and the final interpretations by neuroradiologists. Efforts to reduce residents' errors should focus on the identification of fractures and signs of chronic ischemic change.

Many medical centers offer full-time staffing of the radiology department with board-certified radiologists, while others offer 24-hour CT interpretation either by means of teleradiology or on-call systems. Despite our best efforts, errors in interpretation occur in these circumstances. These problems may be compounded in medical centers where radiology residents provide preliminary interpretations for the emergency department. Because of the relative inexperience of the residents, increased numbers of errors in interpretation are possible. Previous reports have described low (<2%) significant miss rates with residents' interpretations of cranial CT scans obtained to assess trauma (1-4). However, no statistically significant difference in the performances of upper- and lower-level residents has been shown. This study was designed to assess whether more experienced resi-

dents are more accurate in interpreting emergency cranial CT scans than are more junior residents. The types of misses by residents were assessed to help focus educational efforts.

## Methods

Radiology residents ( $n = 18$ ), either during normal working hours or while on call, prospectively interpreted 1324 consecutive head CT scans ordered in the emergency department at the University of Arizona Health Sciences Center, Tucson (a level I trauma center), from January through July 1999. The resident completed and signed a preliminary interpretation form. The form included the resident's interpretation of the head CT image, and a rating of his or her confidence in that interpretation on a six-point scale. The completed form was subsequently faxed to the emergency department.

One of five neuroradiologists with a Certificate of Added Qualification subsequently interpreted the scans, either on the same day or the following morning if the scan was obtained during the on-call shift. The attending neuroradiologist who reviewed the cases was responsible for determining the correlation of the interpretations and for classifying their assessment of the radiologists' interpretations, as follows: agree, disagree-insignificant, disagree-significant. A disagreement (eg, mistaking cytotoxic edema in metastatic disease for ischemic change) was considered significant if an adverse patient outcome was possible or a gross error in synthesis (eg, mistaking dysgenesis of the corpus callosum for chronic hydrocephalus) occurred

---

Received March 2, 2001; accepted after revision August 16.

From the Departments of Radiology (W.K.E., W.G.B., E.K., J.F.S.) and Surgery (J.A.G.), University of Arizona Health Sciences Center, Tucson.

Address reprint requests to William K. Erly, MD, Department of Radiology, University of Arizona, 1501 N. Campbell Avenue, Tucson, AZ 85724-5067.

**TABLE 1: Rates of agreement related to years of training**

Category	Rate of Agreement by Resident Year (%)			
	1st	2nd	3rd	4th
Agree	90	92	94	99
Disagree-insignificant	8	7	6	1
Disagree-significant	2	1	0	0

Note.—The level of training had a significant effect on the rate of agreement ( $P = .032$ ).

without the potential for an adverse outcome. An insignificant error was defined as one in which no potential for an adverse patient outcome (eg, failure to identify ischemic white matter degeneration) existed.

The data were analyzed by using analysis-of-variance methods when one of the variables was numeric (eg, confidence value) or  $\chi^2$  methods when both variables were categorical and when only frequency counts could be obtained from the data (eg, agree, disagree).

A retrospective review of cases involving disagreement was performed to characterize the types of errors and to determine the potential for changes in patient treatment and outcome. A review of patients' charts was performed for cases in which an adverse outcome was possible because of the resident's misinterpretation.

## Results

A total of 770 (58%) of the 1324 scans were considered normal, and 554 (42%) were abnormal. Overall, the agreement rate between the residents' initial interpretations and the attending neuroradiologists' final interpretations was 91% (SD, 8.25); disagree-insignificant rate, 7% (SD, 8.21); and disagree-significant rate, 1.5% (SD, 1.03). No statistically significant differences in rates of agreement or disagreement existed among the neuroradiologists ( $\chi^2 = 13.17$ ,  $df = 8$ ,  $P = 0.11$ ). Disagreements between a resident and an attending physician occurred more often with abnormal than with normal findings ( $\chi^2 = 44.46$ ,  $df = 2$ ,  $P < .001$ ). Of the 90 cases that had insignificant disagreement, 32% involved normal findings, and 68% involved abnormal findings. Of the 23 cases with significant disagreement, 17% involved normal findings, and 83% involved abnormal findings.

### *Effect of Level of Training*

The level of residency training significantly affected rates of agreement ( $\chi^2 = 13.80$ ,  $df = 6$ ,  $P = .032$ ), although these were high for each year of residency. The data are summarized in Table 1. The neuroradiologists agreed with the first-year residents 90% of the time, with 8% disagree-insignificant and 2% disagree-significant rates. They agreed with second-year residents 92% of the time, with 7% disagree-insignificant and 1% disagree-significant rates. The agreement rate with the third-year residents was 94%, with 6% disagree-insignificant rates. The neuroradiologists agreed with the fourth-year residents 99% of the time, with 1% disagree-insignificant rates. Rates of agreement increased with the year of residency train-

ing. All significant disagreements occurred with the first- and second-year residents, although the overall disagree-significant rate was only 2%.

### *Residents' Confidence Levels and Rates of Agreement*

Overall, a significant relationship ( $\chi^2 = 37.55$ ,  $df = 4$ ,  $P < .001$ ) existed between the residents' confidence levels and the agreement or disagreement of their interpretations with those of the attending neuroradiologist. In cases in which agreement occurred, 24% had very confident ratings, 72% had confident ratings, and only 4% had somewhat confident ratings. In cases in which disagreement was insignificant, 16% had very confident ratings, 71% had confident ratings, and 13% had somewhat confident ratings. In the cases with significant disagreement, only 4% of residents were very confident, 78% were confident, and 18% were somewhat confident. In other words, the less confident the resident was in the diagnosis, the more likely it was that some degree of disagreement would occur.

### *Residents' Confidence Levels and Levels of Training*

The confidence that the residents had in their diagnoses differed ( $F = 92.01$ ,  $df = 3$ ,  $P < .001$ ) depending on the year of residency. First-year residents had the lowest confidence ratings; 7% were very confident, 86% were confident, and 7% were somewhat confident. Thirty-four percent of second-year residents were very confident, 65% were confident, and 1% were somewhat confident. Among third-year residents, 62% were very confident, 35% were confident, and 3% were somewhat confident. Thirty-one percent of fourth-year residents were very confident, 68% were confident, and 1% were somewhat confident. Overall, the residents were significantly more confident with normal findings than with abnormal findings ( $\chi^2 = 74.56$ ,  $df = 2$ ,  $P < .001$ ). In the normal findings, 29% were very confident, 70% were confident, and 1% were somewhat confident. In the abnormal cases, only 16% were very confident; 75% were confident, and 9% were somewhat confident. Residents in each year saw approximately the same proportion of cases with normal and abnormal findings.

### *Cases with Insignificant Disagreement*

At retrospective analysis, characterization of the types of insignificant disagreement was possible in 59 of the 90 cases. Forty-two (71%) cases involved false-negative (resident miss) findings, and 17 involved false-positive (resident overcall) findings. The most common resident misses for cases with insignificant disagreement were related to chronic ischemic foci ( $n = 12$ ), fractures ( $n = 9$ ), and atrophy ( $n = 8$ ). The most common resident overcalls in cases with insignificant disagreement involved intracranial hemorrhage ( $n = 8$ ) and fractures ( $n = 6$ ). In the case in

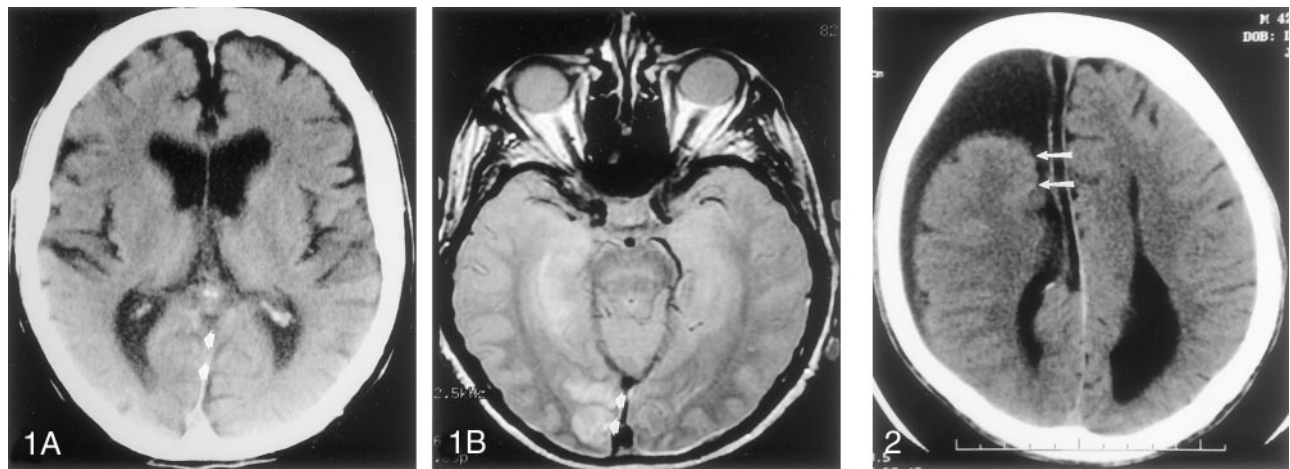


FIG 1. Axial CT scan (A) and MR image (B) show false-negative finding involving ischemic disease.

A, Acute right occipital infarct is visible as both hypoattenuating gray matter and hypoattenuating white matter, with associated sulcal effacement (arrows).

B, Proton density-weighted (2400/30 [TR/TE]) MR image confirms the findings (arrows).

FIG 2. Axial CT image obtained in a patient with schizencephaly and callosal dysgenesis shows an error in synthesis that was considered significant. Note the communication of the right lateral ventricle with the subarachnoid space (arrows) and the characteristic configuration of the occipital horns.

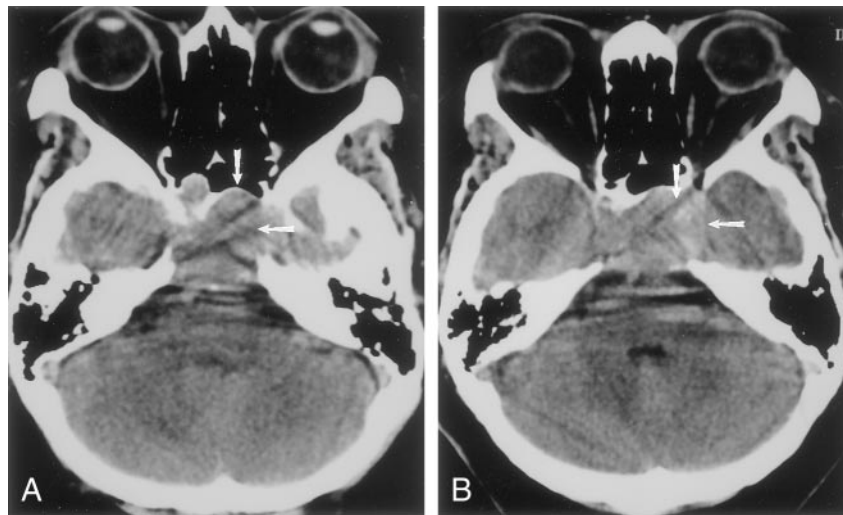


FIG 3. Contiguous 5-mm non-contrast-enhanced routine axial CT scans demonstrate a large mass (arrows) in the sella in a case of pituitary macroadenoma.

which intracranial hemorrhage or fracture was missed, the findings were considered insignificant if they did not alter management. For example, if a subdural hematoma that required surgical evacuation was identified and a small amount of subarachnoid blood was overlooked, the overlooked finding was considered insignificant.

#### Cases with Significant Disagreement

At review of the 23 cases initially classified as disagree-significant, two false-positive interpretations by the attending neuroradiologist were confirmed with negative findings at follow-up study. One case involved a suspected subdural hematoma, and the other involved a suspected contusion. One additional case with significant disagreement involved an initially correct interpretation of the head CT scan by the resident, followed by a misinterpretation of the follow-up

thin-section (3-mm) images, which showed a frontal sinus fracture. When the attending physicians' false-positive interpretations and the one not involving the head CT scan were excluded, 20 (1.5%) cases had significant disagreement.

Fourteen (70%) of these 20 cases involved false-negative interpretations by the resident; nine of these involved facial or skull fractures. Other misses included an acute right occipital infarct (Fig 1); schizencephaly with callosal dysgenesis, which was interpreted as hygroma-hydrocephalus mass effect (Fig 2); a possible temporal lobe contusion (patient lost to follow-up); a chronic cortical infarct; a pituitary macroadenoma (Fig 3); and a small thalamic hemorrhage. All were confirmed at subsequent examination. There were two cases in which the resident misinterpreted the findings of vasogenic edema as ischemia (Fig 4). There were two resident false-positive interpretations (cerebellar hemorrhage, bilateral contusions).

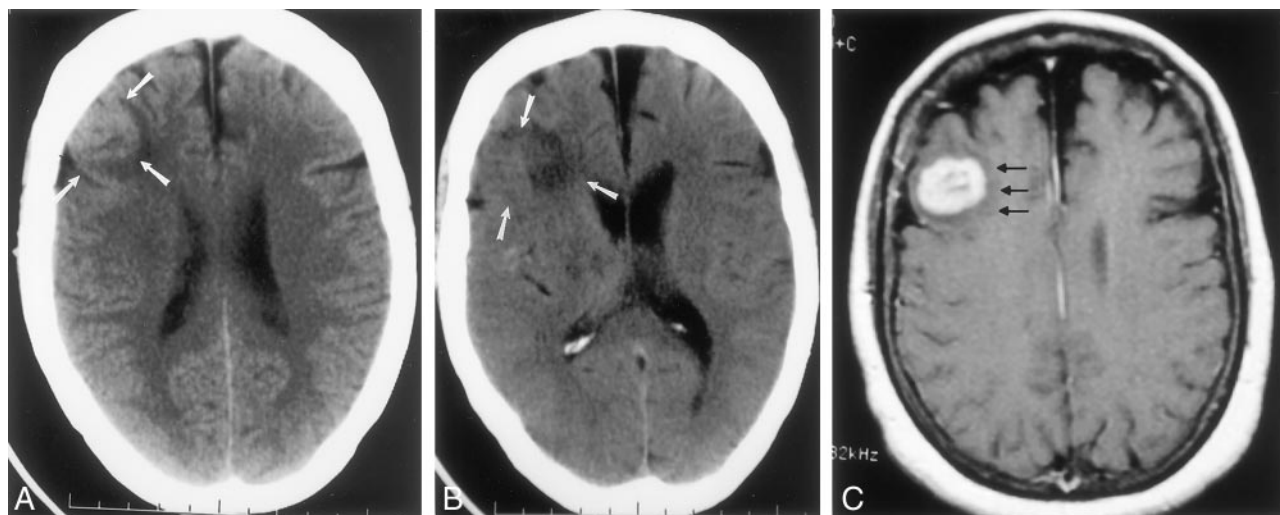


FIG 4. Axial images show metastatic disease interpreted as infarction.  
 A, CT image shows a right frontal lobe mass (arrows).  
 B, On the CT section adjacent to A, vasogenic edema (arrows) is evident.  
 C, Contrast-enhanced MR image more clearly shows the mass (arrows).

TABLE 2: Sources of errors by residents in the interpretation of scans in disagree-significant and disagree-insignificant categories

Type of error	Number of Error	Percentage (%)	
		Of All Scans	Of All Errors
Fracture	18	1.4	16
Chronic ischemic change	12	0.9	11
Atrophy	8	0.6	7
Related to neoplasm	3	0.2	3

Note.—The total scans was 1324. The total number of errors by residents was 113.

Retrospective analysis of the 20 cases with true significant disagreement revealed that patient treatment could have been affected in 16. Review of these cases lead to the identification of seven in which patient outcome could have been adversely affected. Review of these patients' charts revealed that the patient had no specific adverse clinical effect that was related to the resident's misinterpretation.

When both the disagree-significant and disagree-insignificant categories were combined, a clearly identified cause could be assigned in 82 of 113 cases. The main sources of false-negative interpretations by residents are listed in Table 2.

### Specific Diagnoses

Overall, eight cases (0.6%) of acute cerebral ischemia were identified. Of these, the initial interpretation was correct in seven, with one false-negative report and two false-positive reports in which neoplastic disease was reported as ischemic change (sensitivity, 87.5%; specificity, 99.8%). Twenty-nine cases (2.2%) of intracranial hemorrhage were identified, with 27 initially correct interpretations, two false-negative reports, and 10 false-positive reports (sensitivity, 93.1%; specificity, 99.2%). These findings are summarized in Table 3.

TABLE 3: Accuracy of initial interpretations of 1324 emergency cranial CT scans by residents

Diagnosis	Findings			
	True-Positive	False-Positive	False-Negative	True-Negative
Acute cerebral ischemia	7	2	1	1314
Hemorrhage	27	10	2	1285

Note.—For acute cerebral ischemia, specificity and sensitivity were 87.5% and 99.8%, respectively; for hemorrhage, 93.1% and 99.2%, respectively.

### Discussion

The need for early and accurate interpretation of images originating from the emergency department seems, at times, to conflict with the need for radiology residents to gain clinical experience and confidence by working independently. This problem may be particularly true with cranial CT for the assessment of trauma in that many patients can be safely discharged from the emergency department if both neurologic and CT findings are normal (5, 6). Among patients with normal neurologic results, those with abnormal CT scans are more likely to require neurosurgical intervention. Therefore, accurate initial interpretation of cranial CT scans by the resident is imperative. Furthermore, with many hospitals offering 24-hour radiology services by board-certified radiologists, ensuring that on-call residents at academic centers can interpret images at this community standard is important. Substandard care cannot be justified on the basis of the education of residents.

In this study, the overall rate of significant disagreement between the preliminary interpretations by residents and the final interpretations of neuroradiologists was low (2%). Our rates of insignificant and significant disagreement were similar to previously reported data (1–4). The rate of potential change in

patient outcome due to resident misinterpretation also was low (0.05%), and subsequent adverse effects were unlikely to occur.

Previous studies regarding the accuracy of radiology residents in interpreting cranial CT scans have been conducted (1, 3, 4). None of these studies revealed or were conducted to assess statistically significant differences in performance between upper- and lower-level residents. Because residents may be on call as early as 6 months into their training, knowing whether the performance of the relatively inexperienced residents meets the community standard and whether their accuracy substantially improves with training is important. One might expect that greater training and experience would result in more accurate interpretations. Our current findings indicate that although all residents functioned within the standard of care, increased experience resulted in greater accuracy in the interpretation of cranial CT scans.

The overall rate for significant misses was very low at any level of training. This finding agrees with previous ones, which showed that the rate of significant misses by residents is 6.3% (7) in the interpretation of plain radiographs and 1.2% in the interpretation of emergency body CT scans (8). In all of these studies, the final interpretation of the attending physician was used as the criterion standard. Interobserver variability between neuroradiologists in the interpretation of cranial CT scans undoubtedly exists, and in our series, two cases of significant disagreements (9%) were a result of false-positive interpretations by the attending neuroradiologist. Studies of the interpretation of cranial CT scans with teleradiology revealed a 4% rate of significant disagreements (9).

The sensitivity and specificity of the residents' interpretations of acute ischemic changes were 87.5% and 99.8%, respectively, if the final interpretations of the neuroradiologists are assumed to be correct. Other authors have examined the intraobserver variability between neuroradiologists and between emergency physicians, neurologists, and neuroradiologists in their ability to detect changes of early ischemia on CT scans (10, 11). The mean agreement among neuroradiologists is 82% in the identification of the signs of acute stroke (10). Agreement among emergency physicians, stroke specialists, neurologists, and general radiologists is less (11). However, one must be cautious in comparing these numbers to the data acquired in the current study. The prevalence of disease in the sample group may have had a large effect on the interobserver agreement rates. For example, if all of the images in the sample were normal, one would have expected a greater rate of agreement than that expected if all of the CT scans were abnormal.

Although the overall misinterpretation rate was low, errors generally were related to perception rather than to misinterpretation of identified abnormalities. This finding, in part, may be due to the limited scope of the pathologic cases that were referred from the emergency department. In this series, the largest source of errors was missed fractures,

which most commonly involved the face. Less common, and probably less important, were errors related to chronic ischemic disease and atrophy. Although these findings were important, the on-call residents may have dismissed these signs as being unrelated to the patient's acute presentation. The images in two of eight patients with neoplastic disease were initially misinterpreted as showing cerebral infarction. One pituitary macroadenoma was missed. Educational efforts should be made to ensure that the residents check all the bone window images, and signs that can be used to differentiate vasogenic edema from cytotoxic edema should be emphasized.

## Conclusion

The overall rate of significant disagreement between the preliminary resident interpretation and final attending neuroradiologist interpretation of head CT scans ordered in the emergency department is low (2%) and within the community standard. Also, the rate of potential changes in patient outcome due to resident misinterpretations was low (0.05%), and subsequent adverse effects to were unlikely. Efforts to reduce resident errors should focus on the identification of fractures, cytotoxic versus vasogenic edema, and signs of chronic ischemic disease.

## Acknowledgment

The authors wish to thank Kristina Lärka for editorial assistance.

## References

1. Wysoki MG, Nassar CJ, Koenigsberg RA, et al. **Head trauma: CT scan interpretation by radiology residents versus staff radiologists.** *Radiology* 1998;208:125-128
2. Funaki B, Szymiski GX, Rosenblum JD. **Significant on-call misses by radiology residents interpreting computed tomographic studies: perception versus cognition.** *Emerg Radiol* 1997;4:290-294
3. Roszler MH, McCarroll KA, Rashid T, Donovan KR, Kling GA. **Resident interpretation of emergency computed tomographic scans.** *Invest Radiol* 1991;26:374-376
4. Lal NR, Eldevik OP, Murray UM, Desmond JS. **Clinical consequences of misinterpretations of neurologic CT scans by on-call radiology residents.** *AJNR Am J Neuroradiol* 2000;21:124-129
5. Shackford SR, Wald SL, Ross SE, et al. **The clinical utility of computed tomographic scanning and neurologic examination in the management of patients with minor head injuries.** *J Trauma* 1992;33:385-394
6. Stein SC, Ross SE. **Mild head injury: plea for routine early CT scanning.** *J Trauma* 1992;33:11-13
7. Seltzer SE, Hessel SJ, Herman PG, Swenson RG, Sheriff CR. **Resident film interpretations and staff review.** *AJR Am J Roentgenol* 1981;137:129-133
8. Wechsler RJ, Spettell CM, Kurtz AB, et al. **Effects of training and experience in interpretation of emergency body CT scans.** *Radiology* 1996;199:717-720
9. Pysker L, Harlow C. **Teleradiology using low-cost consumer-oriented computer hardware and software.** *AJR Am J Roentgenol* 1998;172:1181-1184
10. von Kummer R, Holle R, Grzyska U, et al. **Interobserver agreement in assessing early CT signs of middle cerebral artery infarction.** *AJNR Am J Neuroradiol* 1996;17:1743-1748
11. Grotta JC, Chiu D, Lu M, et al. **Agreement and variability in the interpretation of early CT changes in stroke patients qualifying for intravenous rtPA therapy.** *Stroke* 1999;30:1528-1533