

Are your **MRI contrast agents** cost-effective?

Learn more about generic **Gadolinium-Based Contrast Agents**.



# AJNR

## **Comparing Preliminary and Final Neuroradiology Reports: What Factors Determine the Differences?**

K. Stankiewicz, M. Cohen, M. Carone, G. Sevinc, P.G. Nagy, J.S. Lewin, D.M. Yousem and L.S. Babiartz

This information is current as of April 20, 2024.

*AJNR Am J Neuroradiol* published online 28 July 2016  
<http://www.ajnr.org/content/early/2016/07/28/ajnr.A4897>

# Comparing Preliminary and Final Neuroradiology Reports: What Factors Determine the Differences?

 K. Stankiewicz,  M. Cohen,  M. Carone,  G. Sevinc,  P.G. Nagy,  J.S. Lewin,  D.M. Yousem, and  L.S. Babiarz

## ABSTRACT

**BACKGROUND AND PURPOSE:** Trainees' interpretations of neuroradiologic studies are finalized by faculty neuroradiologists. We aimed to identify the factors that determine the degree to which the preliminary reports are modified.

**MATERIALS AND METHODS:** The character length of the preliminary and final reports and the percentage character change between the 2 reports were determined for neuroradiology reports composed during November 2012 to October 2013. Examination time, critical finding flag, missed critical finding flag, trainee level, faculty experience, imaging technique, and native-versus-non-native speaker status of the reader were collected. Multivariable linear regression models were used to evaluate the association between mean percentage character change and the various factors.

**RESULTS:** Of 34,661 reports, 2322 (6.7%) were read by radiology residents year 1; 4429 (12.8%), by radiology residents year 2; 3663 (10.6%), by radiology residents year 3; 2249 (6.5%), by radiology residents year 4; and 21,998 (63.5%), by fellows. The overall mean percentage character change was 14.8% (range, 0%–701.8%; median, 6.6%). Mean percentage character change increased for a missed critical finding (+41.6%,  $P < .0001$ ), critical finding flag (+1.8%,  $P < .001$ ), MR imaging studies (+3.6%,  $P < .001$ ), and non-native trainees (+4.2%,  $P = .018$ ). Compared with radiology residents year 1, radiology residents year 2 (−5.4%,  $P = .002$ ), radiology residents year 3 (−5.9%,  $P = .002$ ), radiology residents year 4 (−8.2%,  $P < .001$ ), and fellows (−8.7%,  $P < .001$ ) had a decreased mean percentage character change. Senior faculty had a lower mean percentage character change (−6.88%,  $P < .001$ ). Examination time and non-native faculty did not affect mean percentage character change.

**CONCLUSIONS:** A missed critical finding, critical finding flag, MR imaging technique, trainee level, faculty experience level, and non-native-trainee status are associated with a higher degree of modification of a preliminary report. Understanding the factors that influence the extent of report revisions could improve the quality of report generation and trainee education.

**ABBREVIATIONS:** CF = critical finding; M = modified; PCC = percentage character change; R1 = radiology resident year 1; R2 = radiology resident year 2; R3 = radiology resident year 3; R4 = radiology resident year 4

Understanding the prevalence, causes, and types of discrepancies and errors in examination interpretation is a critical step in improving the quality of radiology reports. In an academic setting, discrepancies and errors can result from nonuniform training levels of residents and fellows. However, even the “ex-

perts” err, and a prior study found a 2.0% clinically significant discrepancy rate among academic neuroradiologists.<sup>1</sup> A number of factors can affect the accuracy of radiology reports. One variable of interest at teaching hospitals is the effect of the involvement of trainees on discrepancies in radiology reports. Researchers have found that compared with studies read by faculty alone, the rate of clinically significant detection or interpretation error was 26% higher when studies were initially reviewed by residents, and it was 8% lower when the studies were initially interpreted by fellows.<sup>2</sup> These findings suggest that perhaps faculty placed too much trust in resident interpretations, which led to a higher rate of discrepancies, while on the other hand, having a second experienced neuroradiology fellow look at a case can help in reducing the error rate.<sup>2</sup>

In our academic setting, preliminary reports initially created by trainees are subsequently reviewed and finalized by faculty or staff. The changes made to preliminary reports are a valuable teaching tool for trainees because clear and accurate report writ-

Received February 18, 2016; accepted after revision May 16.

From The Russell H. Morgan Department of Radiology and Radiological Sciences (K.S., M. Cohen, G.S., P.G.N., J.S.L., D.M.Y., L.S.B.), Johns Hopkins Medical Institutions, Baltimore, Maryland; and Department of Biostatistics, University of Washington (M. Carone), Seattle, Washington.

Paper previously presented at: Annual Meeting of the American Society of Neuroradiology and the Foundation of the ASNR Symposium, May 17–22, 2014; Montreal, Quebec, Canada.

Please address correspondence to David M. Yousem, MD, The Russell H. Morgan Department of Radiology and Radiological Sciences, Johns Hopkins Medical Institutions, 600 N Wolfe St, Phipps B-100F, Baltimore, MD 21287; e-mail: dyousem1@jhu.edu

<http://dx.doi.org/10.3174/ajnr.A4897>

ing is a critical skill for a radiologist.<sup>3</sup> Recently, computer-based tools have been created to help trainees compare the changes between preliminary and final reports to improve their clinical skills and to facilitate their learning. Sharpe et al<sup>4</sup> described the implementation of a Radiology Report Comparator, which allows trainees to view a merged preliminary/final report with all the insertions and deletions highlighted in “tracking” mode. Surrey et al<sup>5</sup> proposed using the Levenshtein percentage or percentage character change (PCC) between preliminary and final reports as a quantitative method of indirectly assessing the quality of preliminary reports and trainee performance. The Levenshtein percentage, a metric used in computer science, compares 2 texts by calculating the total number of single-character changes between the 2 documents, divided by the total character count in the final text.<sup>5</sup>

In this study, we analyzed preliminary neuroradiology reports dictated by trainees and the subsequent finalized reports revised by our faculty. We set out to identify the factors that determine the degree to which the preliminary reports are modified by faculty for residents and fellows, for daytime and nighttime shifts, and for CT and MR imaging examinations. We hypothesized that study complexity, lack of experience (for both trainee and faculty), and perhaps limited language skills (native-versus-non-native speaker) would result in a greater number of corrections.

## MATERIALS AND METHODS

In accordance with the Health Insurance Portability and Accountability Act, our institutional review board reviewed and approved the protocol for this retrospective study and waived the requirement for informed consent.

### Study Sample

Using our electronic medical records and Radiology Information System, we identified all neuroradiology reports generated at our institution between November 1, 2012, and October 31, 2013 (12 consecutive months). Neuroradiology reports by faculty only were excluded. At our institution, 80% of all neuroradiology studies are interpreted by trainees and faculty, and 20% are interpreted by faculty alone. Similar to individuals at other academic medical centers, our trainees, residents, and fellows create preliminary reports that are subsequently reviewed and, if necessary, revised by our faculty. Our entire faculty is neuroradiology fellowship-trained. Because preliminary reports are released into the electronic medical records and are viewable by the referring clinicians, if a significant change is made to the preliminary report, the final report is marked with an electronic flag (M), for modified. At our institution, the ordering or current provider is not automatically alerted to the change, but rather, our faculty or trainee (after discussing the changes with the faculty) communicates directly with the primary clinical team.

Per recommendation of the American College of Radiology and The Joint Commission, our trainees and faculty verbally communicate with the primary clinical team about neuroradiologic abnormal findings that may have immediate impact on patient care. At our institution, a predetermined list of 17 critical findings has been developed, which includes new hemorrhage, new stroke, new/increasing mass, increasing intracranial pressure, new/worsening herniation, new/worsening hydrocephalus,

misplaced/malfunctioning surgical hardware, infection, child abuse, vascular abnormality, new cord compression, new cord infarction, new spinal instability, congenital variations altering surgical approach, acute fracture, and globe/retina/optic nerve compromise. All neuroradiology reports containing a critical finding (CF) are electronically marked with a Flag (C) for ease of identification and documentation.

We have 10 residents and 9 neuroradiology fellows per year. Each year on July 1, the trainee graduates to a higher residency level or fellowship. Because our 12 consecutive months of reports encompasses that transition date, the same trainee may have been designated as a radiology resident year 1 (R1), year 2 (R2), year 3 (R3), or year 4 (R4) and neuroradiology fellow, depending on when the examination was performed.

Our faculty on staff was subdivided into junior, intermediate, and senior faculty based on <3, 3 years but <7 years, and >7 years of experience in practice after fellowship.

For each trainee and faculty member, the native-versus-non-native English-speaker status was recorded. We defined a non-English speaker as an individual who did not enter an English-speaking educational system until high school.

For examinations performed at the same time and involving consecutive body parts, such as CT of the cervical, thoracic, and lumbar spine or CT of the head and maxillofacial region, frequently, our trainee and faculty member dictate a single report, which is then attached to each individual study accession number. We avoided analyzing duplicate reports by including only the reports that had images attached to them. In our Radiology Information System, a complete set of images from an examination can only be attached to a single report, regardless of how many accession numbers (Current Procedural Terminology codes) are linked to that report.

At our institution, trainees are under direct faculty supervision during daytime hours (7 AM to 11 PM Monday through Friday, and 8 AM to 11 PM on Saturday and Sunday). During that time, a faculty member is always available for consultation. During nighttime hours (11 PM to 7 AM Monday through Friday, and 11 PM to 8 AM Saturday and Sunday), trainees interpret studies more independently; however, they can use our paging system to contact a faculty member for consultation. The “examination end” time stamp was used to determine whether a study was performed during daytime or nighttime.

At our institution, we do not use report templates for neuroradiology examinations. This choice likely increases the variability among our reports and has an impact on the extent of revisions to the preliminary reports performed by our faculty.

In an automated fashion, the percentage character change between the preliminary report generated by the trainee and the final report revised and signed by faculty was determined. The character change was defined as the total number of single-character changes between the preliminary and final report. The percentage character change was defined as  $PCC = (100 \times \text{Total Number of Single Character Changes}) / (\text{Total Character Number in Original Report})$ . Because the total number of single-character changes can exceed the number of characters in the original report, this PCC value can be any non-negative percentage, even exceeding 100%.

### Distribution of neuroradiology reports by percentage character change

PCC	No. of Reports	Percentage of Total Reports	95% CI
0 (no change)	95	0.3%	0.2%–0.5%
0% < x ≤ 1%	10,892	31.4%	28.5%–34.5%
1% < x ≤ 5%	4888	14.1%	13.0%–15.2%
5% < x ≤ 10%	4232	12.2%	11.4%–13.0%
10% < x ≤ 25%	7351	21.2%	20.0%–22.5%
25% < x ≤ 50%	4918	14.2%	12.8%–15.7%
50% < x ≤ 100%	2007	5.8%	5.0%–6.8%
>100%	278	0.8%	0.6%–1.0%

### Statistical Analysis

Basic descriptive statistics were calculated to characterize the various key features of the preliminary and final reports. A multivariable linear regression model was used to evaluate the joint association between mean PCC and each of a variety of factors including the following: 1) the presence of a critical or missed finding, 2) whether the report was written during nighttime or daytime, 3) imaging technique (CT or MR imaging), 4) English language proficiency of both the trainee and the faculty, and 5) the seniority of both the trainee and the faculty. Point estimates and confidence intervals for model parameters were obtained using generalized estimating equations with a working independence correlation matrix and robust variance estimators to appropriately account for the possible correlation between reports involving the same trainee and faculty. Generalized estimating equations were also used to provide valid confidence intervals for the marginal PCC value distribution. Plots of model residuals by either attending or trainee were scrutinized to determine whether reports written or edited by any attending or trainee had a substantially greater mean PCC than predicted by the fitted model, adjusting for all factors listed above. This procedure allowed us to scrutinize whether results reported were driven primarily by one or several anomalous individuals. All hypothesis tests were 2-sided and conducted at a significance level of .05. All computations were performed by using the R statistical programming language.<sup>6</sup>

### RESULTS

In this study, 34,661 sets of preliminary/final reports were included. The mean PCC of all reports was 14.8%, with a minimum of 0%, a maximum of 701.2%, and a median of 6.6%. The distribution of reports by PCC is shown in the Table. Ninety-five reports had a PCC of 0%, indicating that there were no changes between the preliminary and the final reports.

Of all studies, 21,204 (61.2%) were CTs (with an average final character count of 1921.3) and 13,457 (38.8%) were MRIs (with average final character count of 2616.4). After we adjusted for the presence of a CF flag, missed finding, examination time, non-native-speaker status, and experience levels, the mean PCC for MR imaging reports was greater than that for CT reports by 3.6 percentage points (95% CI, +2.5 to +4.8%;  $P < .001$ ).

Of all examinations, 21,998 (63.5%) were interpreted by fellows; 2322 (6.7%), by first-year radiology residents; 4429 (12.8%), by second-year radiology residents; 3663 (10.6%), by third-year radiology residents (R3s); and 2249 (6.5%) by fourth-year radiology residents. When reports created by each category of trainee were compared with a baseline of the reports generated by

R1s, we found that after we adjusted for presence of a CF flag, missed finding, examination time, imaging technique, non-native-speaker status, and faculty experience level, reports created by R2s had a mean PCC lower by 5.4 percentage points (95% CI, –8.8% to –2.0%;  $P = .002$ ), reports created by R3s had a mean PCC lower by 5.9 percentage points (95% CI, –9.5% to –2.2%;  $P = .002$ ), reports created by R4s had a mean PCC lower by 8.2 percentage points (95% CI, –12.5% to –3.8%;  $P < .001$ ), and reports created by fellows had a mean PCC lower by 8.7 percentage points (95% CI, –12.2% to –5.2%;  $P < .001$ ).

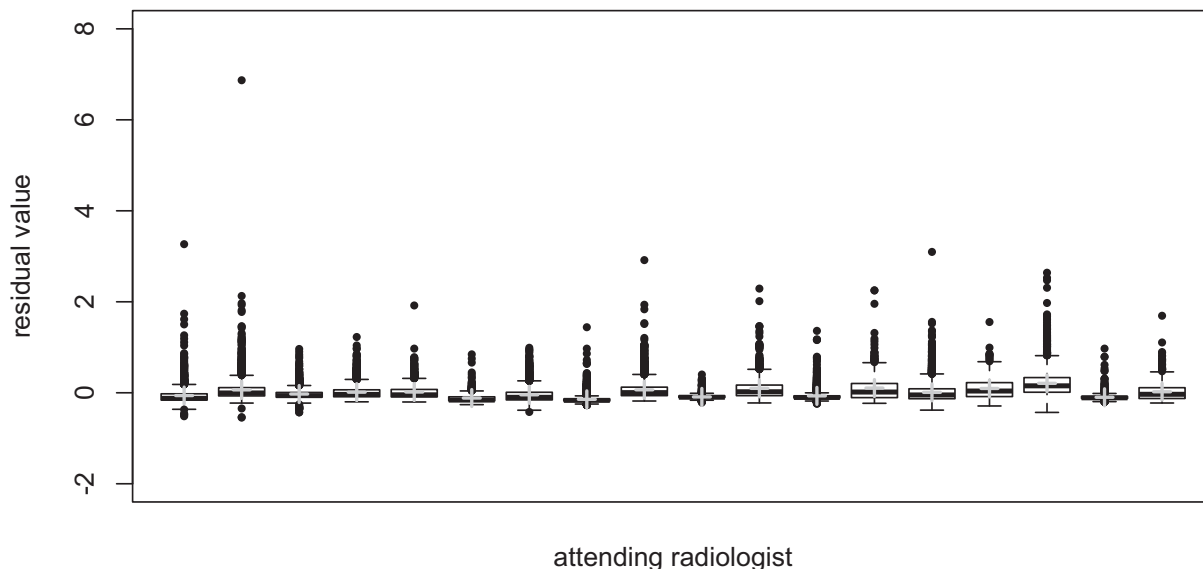
The distribution of cases read by fellows and R1–4 are a product of our neuroradiology rotation and call schedules, with most neuroradiology cases being read by our fellows and R2s and R3s. Our R4s typically take electives related to their planned fellowship; therefore, the few R4s in our division end up staying for a neuroradiology fellowship.

Twelve of the 58 (20.6%) trainees were non-native English speakers, and they accounted for 8808 (25.4%) of all preliminary reports. After we adjusted for the presence of a CF flag, missed finding, examination time, imaging technique, faculty non-native-speaker status, and seniority, these reports had a mean PCC higher by 4.2 percentage points compared with those generated by the native-speaker trainees (95% CI, +0.7% to +7.6%;  $P = .018$ ).

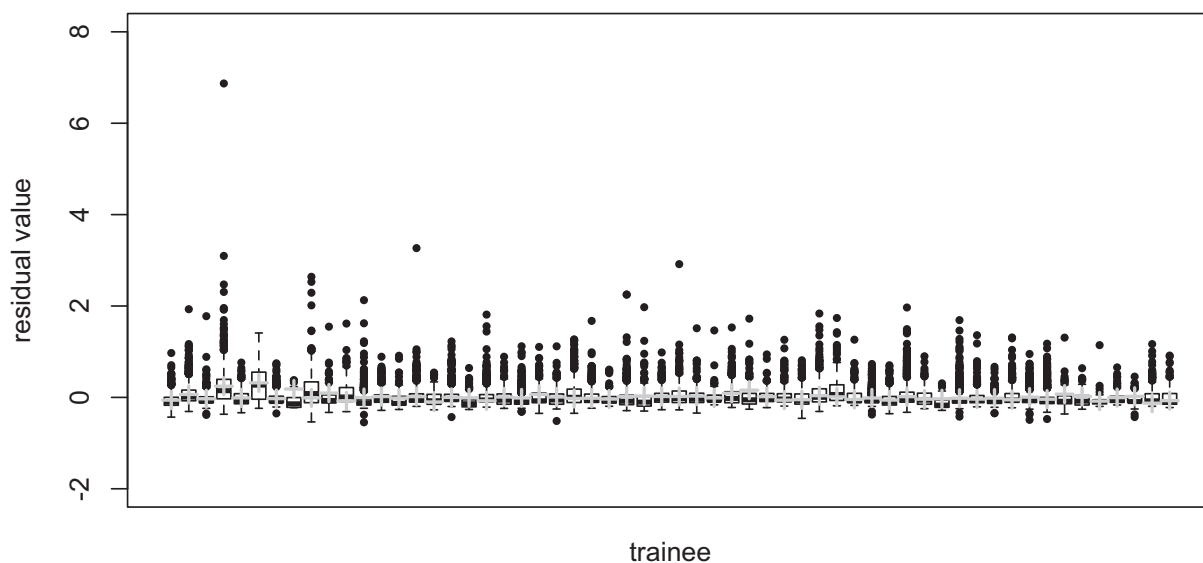
Of all reports, 4091 (11.8%) were marked with a critical finding flag, and 282 of these (6.9% of reports with a critical finding, 0.8% of all reports) were marked with a missed finding flag. After we adjusted for examination time, imaging technique, non-native-speaker status, and experience levels, reports with a CF flag but no missed finding had a mean PCC higher by 1.8 percentage points compared with those without any CF flag (95% CI, +0.9% to +2.7%;  $P < .001$ ), while reports with flags for both a critical finding and missed finding had a mean PCC higher by 41.6 percentage points compared with those with only the CF flag (95% CI, +37.3 to +48.9%;  $P < .001$ ).

Of all reports, 20,123 (58.1%) were created during daytime shifts (under direct faculty supervision), and 14,538 (41.9%), during nighttime (no direct supervision; however, faculty were available for consultation via the paging system). No significant difference was detected between the mean PCC of daytime and nighttime reports (mean PCC higher during nighttime by 0.6 percentage points; 95% CI, –1.4% to +2.6%;  $P = .567$ ) after adjusting for the presence of a CF flag, imaging technique, non-native-speaker status, and experience levels. However, in our sample, the odds of finding a flag M in reports written at night were >2 times higher (OR estimate, 2.02; 95% CI, 1.48–2.77;  $P < .001$ ) than the odds of finding a flag M in reports written during the day.

Eighteen staff neuroradiologists reviewed preliminary reports generated by trainees. Among faculty, there were 4 junior attending neuroradiologists (3920 reports), 5 intermediate attending neuroradiologists (12,885 reports), and 9 senior attending neuroradiologists (17,856 reports). Compared with a baseline of reports finalized by junior neuroradiologists and adjusting for the presence of a CF flag, missed finding, examination time, imaging technique, non-native-speaker status, and trainee experience level, reports finalized by intermediate neuroradiologists showed no significant difference in mean PCC (lower for intermediate neuroradiologists by 0.7 percentage points; 95% CI, –4.2% to



**FIG 1.** Plot of model residuals by attending radiologists. For each attending radiologist, a boxplot of residuals is shown in black and the average residual value is depicted in gray.



**FIG 2.** Plot of model residuals by trainees. For each trainee, a boxplot of residuals is shown in black and the average residual value is depicted in gray.

+2.9%;  $P = .720$ ), while reports finalized by senior neuroradiologists had a mean PCC lower by 6.8 percentage points (95% CI,  $-9.4$  to  $-4.2\%$ ;  $P < .001$ ).

Ten of 18 faculty (55.5%) were non-native English speakers, and they finalized 23,058 (66.5%) of all reports. After adjusting for the presence of a CF flag, missed finding, examination time, imaging technique, trainee non-native-speaker status, and experience levels, we detected no significant difference in the mean PCC for reports finalized by native and non-native-speaker faculty (mean PCC for non-native-speaker faculty higher by 1.1 percentage points; 95% CI,  $-0.8\%$  to  $+3.0\%$ ,  $P = .241$ ).

Plots of model residuals by attendings (Fig 1) did not identify any faculty who, on average, made a greater number of changes than predicted by the model. Plots of model residuals by trainees (Fig 2) also did not reveal any trainee who, on average, had a much

greater number of changes made to his or her reports than predicted by the model.

## DISCUSSION

Overall, in our sample, the mean PCC values were lower in trainees with greater seniority and experience. Reports created by R1s had the highest PCC, and reports created by fellows had the lowest PCC (lower by 8.9 percentage points compared with those generated by R1s). This finding supports our hypothesis that trainees learn to write higher quality reports during their training. The factor with the strongest association to mean PCC was the presence of a flag M or, in other words, a missed CF by a trainee, which, on average, increased the mean PCC by 41.6 percentage points. Even when not missed, the presence of a CF was associated with an increased mean PCC of 1.8 percentage points. Studies

with a CF typically contain more complex pathology; thus, their interpretations are more challenging. This difference increases the potential for error. Additionally, because the flagged studies, particularly ones with a flag M, may have greater implications for patient care, faculty may be more attentive to revising those reports, to ensure that all the findings are accurate and described with precise language. Reports for MR imaging examinations had a higher mean PCC by 3.7 percentage points compared with those for CT examinations. This is most likely caused by the increased complexity of MR imaging studies, which tend to have longer reports, contain more information, and can be more challenging to interpret, especially given the wide range of sequences and protocols. In addition, frequently, MR imaging is used in more complicated cases, increasing the probability that trainees may be exposed to unfamiliar imaging findings and disease processes.

We found a mean PCC higher by 4.0 percentage points in preliminary reports created by trainees who are non-native speakers. This may result from a range of stylistic and vocabulary differences among these trainees, as well as between the non-native-speaking trainees and native-speaking faculty, which could increase the extent of changes made to these reports. No statistically significant difference in the mean PCC was observed between reports finalized by native and non-native English-speaking faculty. This finding suggests that with extensive training and experience in neuroradiology, native and non-native English-speaking faculty adopt similar dictation styles. Few studies have looked at the native and non-native-speaker status of trainees in programs in the United States and its effect on the quality of radiology reports. One potential source of the difference in the mean PCC may be related to the accuracy of the voice-recognition system used by the trainees. Reports generated by non-native English speakers with accents using voice recognition have been shown to have higher error rates of approximately 11.6%, compared with 9.7% for native speakers.<sup>7</sup>

While no statistically significant difference in mean PCC was detected between reports finalized by junior and intermediate faculty, reports finalized by senior faculty exhibited a mean PCC lower by 6.9 percentage points. We hypothesize that junior faculty with limited supervisory experience may be less comfortable with alternate phrasing; thus, they make more changes when editing reports. After we adjusted for the presence of a CF flag, imaging technique, non-native-speaker status, and experience levels, we did not find a statistically significant difference in the mean PCC between studies read during the daytime (under direct supervision) and during the nighttime (without direct supervision, but with faculty available through the paging system). However, in our sample, the odds of finding a flag M in reports written without direct supervision were 2 times higher than the odds of finding a flag M in reports written under direct supervision. One potential explanation for the increased frequency of flag M's is that while supervised, trainees are more likely to consult with their attending neuroradiologist about challenging cases or findings of which they are unsure; thus, such preliminary reports have a lower potential for errors.

Previous studies have shown that the mean PCC values of subsequent sets of preliminary and final reports written by individual trainees exhibit a decreasing trend as trainees advance through radiology training. Sharpe et al<sup>8</sup> studied the average PCC of 6 trainees during their diagnostic radiology residency and

found similar trends among all of them, with the mean PCC falling from 15%–30% to below 15% after about 700 written reports. In our study, we have found a similar trend among all trainees in a large academic hospital because PCC values were lower in each consecutive year of residency and were lowest for fellows.

Our study found a mean PCC of 14.8% and a median PCC of 6.6%. Surrey et al<sup>5</sup> reported a mean value of 6.38%. A few likely factors caused this difference. Surrey et al reported no change between preliminary and final reports in 56.2% of report pairs. In our study, we observed no change in just 95 of the total 34,661 report pairs (~0.3%). This is most likely indicative of more conservative editing on the part of faculty in that study,<sup>5</sup> which would explain the lower mean PCC found. Our hospital does not use templates for radiologic reporting. Institutions using such templates would most likely report lower PCC values because the use of such templates increases conformity among reports and therefore may decrease the proportion of changes made by attending faculty. Although whether such templates were used in the study by Surrey et al is unknown, this is another factor potentially contributing to the differences in mean PCC between the 2 studies.

The influence of direct supervision of trainees by radiology faculty has been a subject of extensive scrutiny. Although trainees do not report any difference in educational value when working with and without direct supervision,<sup>9</sup> previous studies have found that interpretations done by unsupervised trainees had higher discrepancy rates, particularly among less experienced trainees (17% higher discrepancy rates for R2s compared with approximately 7.5% for R3s and R4s and 3.5% for fellows).<sup>2</sup> In a separate review of 18,185 studies interpreted by trainees without supervision, 28 cases of trainee discrepancy later caught by an attending radiologist were estimated to lead to increased morbidity in 11% of the cases and prolonged hospitalization in 25% of the cases, but no case exhibited implications for long-term patient health.<sup>10</sup> There is no consensus, however, with another study finding that just 0.3% of all discrepancies attributed to trainees having no direct supervision resulted in significant negative effects for patients.<sup>11</sup> We did not detect any significant difference in mean PCC between preliminary reports made under direct supervision (daytime) and indirect supervision (nighttime). This can be explained in a few ways: Either the paging system is a sufficient substitute for trainee-faculty consultation or trainees are more attentive while working without attending supervision and are able to largely offset the experience gap.

In this study, we found that the mean PCC was higher in MR imaging than in CT studies, and we hypothesize that this increase is due to a larger amount of discrepant readings and higher report complexity. Indeed, in a study of 416,413 studies read by trainees and reviewed by faculty, researchers found that the discrepancy rate was significantly higher for MR imaging (3.7%) than for CT (2.4%).<sup>12</sup> This same study also found that total discrepancy rates decreased as trainees gained experience, from 1.8% to 1.5%.

Previous applications of the PCC have largely focused on studying individual trainees, not inspecting wider trends in reporting. Researchers suggested using the PCC to identify trainees who may need increased individualized attention and to track the development of trainee reporting skills over the duration of their education. In our study, we assessed the influence of several factors on the PCC. To our knowledge, many of them, such as the

time of the study, native-speaking status, or critical finding flag, have not been examined in this manner. The primary goal of our study was to qualitatively assess which of these factors led to significant changes in the PCC (which had been shown to correlate inversely with report quality). A secondary goal was to achieve a relative quantitative estimate of the magnitude of the aforementioned effects. With our study, we were hoping to identify specific trends that could be targeted with educational effort to improve the quality of our preliminary reports.

The mean PCC itself as a measure of the clinical accuracy of a report has limitations because it weighs all changes equally. For example, reports with a few critical edits (eg, “no stroke” to “stroke”) could have low a PCC but large implications for patient care, and reports with extensive changes (eg, more detailed descriptions, secondary findings) could have a high PCC with little to no implications for patient care. In this study, we did not analyze the content of reports to determine to what extent the mean PCC measures stylistic changes versus meaning of a report. However, another study, performed at our institution by Huntley et al,<sup>13</sup> looking at all neuroradiology reports with a flag M during a 2-year period, did reveal that 73.8% of reports had addenda because of a missed CF, 21.7% had addenda because of a missed non-CF, and 4.6% had addenda because a report was changed from containing a CF to not containing a CF.

There are several limitations to this study. Most important, our study considered only reports generated during 12 consecutive months at 1 university hospital. This means that for example, the 2322 reports written by RIs were written by just 15 residents. With this low sample size, individual trends among trainees and faculty members can significantly influence the results of the entire group. In addition, our RIs come from various training paths, with some having greater experience and knowledge of radiology than others. Indeed, this sort of influence due to individual faculty members has been suggested in previous studies involving PCC.<sup>14</sup> Although in our sample, we did not identify any faculty or trainee outliers, repeating similar experiments at different hospitals and across time can help ensure the precision of our results. Also, while previous studies have shown that in large datasets, the PCC correlates with the clinical accuracy of the report,<sup>5</sup> to our knowledge, no studies have been performed to quantify the magnitude to which other factors, such as changes to formatting, grammar, or spelling, influence the PCC values in radiology. This subject is of particular importance when measuring the impact of variables such as non-native English-speaking status because in this study, we hypothesized that these variables affect the PCC significantly. At our institution, we do not use templates/structured reporting for neuroradiology studies. This feature likely increases the variability between reports and the extent of revisions to the preliminary reports.

## CONCLUSIONS

Our analysis showed that having a CF in the report, missing a finding, MR imaging technique, trainee and faculty experience levels, and non-native-speaker trainee status are associated with a higher degree of modification of a preliminary neuroradiology report. Understanding the factors that influence the extent of report revisions could improve the quality of report generation and trainee education.

Disclosures: Gorkem Sevinc—UNRELATED: Other: I have board memberships in Insight Medical Technologies and emocha Mobile Health. These companies are not relevant to this article. Insight Medical technologies provides expert medical consults to attorneys, and emocha Mobile Health is a mobile health platform company in the public health space. Jonathan S. Lewin—UNRELATED: Board Membership: American Roentgen Ray Society, Association of University Radiologists, Society of Chairs of Academic Radiology Departments, International Society for Strategic Studies in Radiology, Comments: no payment for any of the boards mentioned; Travel/Accommodations/Meeting Expenses Unrelated to Activities Listed: reimbursements from several of the societies noted above for board or committee meetings. David M. Yousem—UNRELATED: Consultancy: medicolegal consultations; Expert Testimony: medicolegal testimony; Payment for Lectures (including service on Speakers Bureaus): American College of Radiology Education Center\*; Royalties: 4 books published by Elsevier; Payment for Development of Educational Presentations: CMEInfo.com\*; Travel/Accommodations/Meeting Expenses Unrelated to Activities Listed: American Society for Neuroradiology international program. \*Money paid to the institution.

## REFERENCES

- Babiarz LS, Yousem DM. **Quality control in neuroradiology: discrepancies in image interpretation among academic neuroradiologists.** *AJNR Am J Neuroradiol* 2012;33:37–42 CrossRef Medline
- Viertel VG, Babiarz LS, Carone M, et al. **Quality control in neuroradiology: impact of trainees on discrepancy rates.** *AJNR Am J Neuroradiol* 2012;33:1032–36 CrossRef Medline
- ACGME Program Requirements for Graduate Medical Education in Diagnostic Radiology. [https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/420\\_diagnostic\\_radiology\\_2016.pdf](https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/420_diagnostic_radiology_2016.pdf). Effective July 1, 2016. Accessed July 14, 2016
- Sharpe RE Jr, Surrey D, Gorniak RJ, et al. **Radiology report comparator: a novel method to augment resident education.** *J Digit Imaging* 2012;25:330–36 CrossRef Medline
- Surrey D, Sharpe RE Jr, Gorniak RJ, et al. **QRSE: a novel metric for the evaluation of trainee radiologist reporting skills.** *J Digit Imaging* 2013;26:678–82 CrossRef Medline
- R. The R Project for Statistical Computing. <http://www.R-project.org/>. Accessed May 20, 2016
- Kanal K, Hangiandreou N, Sykes A, et al. **Initial evaluation of a continuous speech recognition program for radiology.** *J Digit Imaging* 2001;14:30–37 CrossRef Medline
- Sharpe RE, Pan K, Gorniak RJ, et al. **Tracking reporting metrics over the course of a resident/fellow training period: an automated methodology to evaluate progress in the reporting process.** In: *Proceedings of the Annual Meeting of the Society for Imaging Informatics*, Dallas, Texas. June 6–9, 2013
- McGrath JR, Mitra S, Nicola R. **The effect of overnight attending coverage on resident education at a large academic medical center.** In: *Proceedings of the Annual Meeting of the Association of University Radiologists*, Baltimore, Maryland. April 1–4, 2014
- Friedman SM, Merman E, Chopra A. **Clinical impact of diagnostic imaging discrepancy by radiology trainees in an urban teaching hospital emergency department.** *Int J Emerg Med* 2013;6:24 CrossRef Medline
- Ruchman RB, Jaeger J, Wiggins EF 3rd, et al. **Preliminary radiology resident interpretations versus final attending radiologist interpretations and the impact on patient care in a community hospital.** *AJR Am J Roentgenol* 2007;189:523–26 CrossRef Medline
- Weinberg BD, Richter MD, Champine JG, et al. **Radiology resident preliminary reporting in an independent call environment: multi-year assessment of volume, timeliness, and accuracy.** *J Am Coll Radiol* 2015;12:95–100 CrossRef Medline
- Huntley JH, Carone M, Yousem DM, et al. **Opportunities for targeted education: critical neuroradiologic findings missed or misinterpreted by residents and fellows.** *AJR Am J Roentgenol* 2015;205:1155–59 CrossRef Medline
- Surrey D, Sharpe RE, Gorniak RJ, et al. **Variability in the extent of edits made by attending radiologists to trainee radiology reports.** In: *Proceedings of the Annual Meeting of the Association of University Radiologists*, San Antonio, Texas. March 19–22, 2012