AJNR

Am I Ready to Be an Independent Neuroradiologist? Objective Trends in Neuroradiology Fellows' Performance during the Fellowship Year

J.H. Masur, J.E. Schmitt, D. Lalevic, T.S. Cook, L.J. Bagley, S. Mohan and A.P. Nayate

AJNR Am J Neuroradiol 2021, 42 (5) 815-823
doi: https://doi.org/10.3174/ajnr.A7030
http://www.ajnr.org/content/42/5/815

Am I Ready to Be an Independent Neuroradiologist? Objective Trends in Neuroradiology Fellows' Performance during the Fellowship Year

J.H. Masur, J.E. Schmitt, D. Lalevic, T.S. Cook, L.J. Bagley, S. Mohan, and A.P. Nayate

ABSTRACT

BACKGROUND AND PURPOSE: Aside from basic Accreditation Council for Graduate Medical Education guidelines, few metrics are in place to monitor fellows' progress. The purpose of this study was to determine objective trends in neuroradiology fellowship training on-call performance during an academic year.

MATERIALS AND METHODS: We retrospectively reviewed the number of cross-sectional neuroimaging studies dictated with complete reports by neuroradiology fellows during independent call. Monthly trends in total call cases, report turnaround times, relationships between volume and report turnaround times, and words addended to preliminary reports by attending neuroradiologists were evaluated with regression models. Monthly variation in frequencies of call-discrepancy macros were assessed via χ^2 tests. Changes in frequencies of specific macro use between fellowship semesters were assessed via serial 2-sample tests of proportions.

RESULTS: From 2012 to 2017, for 29 fellows, monthly median report turnaround times significantly decreased during the academic year: July (first month) = 79 minutes (95% CI, 71–86 minutes) and June (12th month) = 55 minutes (95% CI, 52–60 minutes; *P* value = .023). Monthly report turnaround times were inversely correlated with total volumes for CT ($r = -0.70$, $F = 9.639$, *P* value = .011) but not MR imaging. Words addended to preliminary reports, a surrogate measurement of report clarity, slightly improved and discrepancy rates decreased during the last 6 months of fellowship. A nadir for report turnaround times, discrepancy errors, and words addended to reports was seen in December and January.

CONCLUSIONS: Progress through fellowship correlates with a decline in report turnaround times and discrepancy rates for cross-sectional neuroimaging call studies and slight improvement in indirect quantitative measurement of report clarity. These metrics can be tracked throughout the academic year, and the midyear would be a logical time point for programs to assess objective progress of fellows and address any deficiencies.

ABBREVIATIONS: CSNI = cross-sectional neuroimaging studies; RIS = radiology information system; RTAT = report turnaround time

A fellow's progress in an academic year is primarily assessed using qualitative, thus subjective, criteria, including achievement of Accreditation Council for Graduate Medical Education–prescribed milestones and faculty evaluations. While the Accreditation Council for Graduate Medical Education provides requirements for total yearly cases read[1] and individual programs may have internal metrics for fellows' progress, there are no concrete external objective measurements for documenting fellows' progress within the academic year. Often, fellows are unsure whether their efficiency in generating reports, report turnaround times (RTATs) for on-call examinations, or quality of on-call reports is satisfactory.

The total number of studies dictated by the fellow and the RTATs of on-call studies may be reviewed by the attendings and program director with the fellows, but more meaningful interpretation of these numbers is lacking because there are no comparison benchmarks or quantitative checkpoints within the fellowship year. Knowledge of these factors is critical in a fellowship program so that program directors and fellows are jointly aware of progress throughout the year and remediation or additional focused training can be implemented, as necessary. More data on neuroradiology fellowship training are especially needed because a survey in 2016

Received May 31, 2020; accepted after revision November 19.

From the Department of Radiology (J.H.M., J.E.S., D.L., T.S.C., L.J.B., S.M.), Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania; and Department of Radiology (A.P.N.), University Hospitals Cleveland Medical Center, Cleveland, Ohio.

J.H. Masur and J.E. Schmitt are co-first authors.

Paper previously presented, in part, at: Annual Meeting of the American Society of Neuroradiology, May 18-23, 2019; Boston, Massachusetts.

Please address correspondence to Ameya P. Nayate, MD, Department of Radiology, University Hospitals Cleveland Medical Center, 11100 Euclid Ave, BSH 5056, Cleveland, OH 44106; e-mail: Ameya.Nayate@UHhospitals.org

http://dx.doi.org/10.3174/ajnr.A7030

demonstrated that 25% of practicing neuroradiologists in the United States believe that fellows' abilities have declined.[2] Prior studies have analyzed various other factors related to radiology residency training, including total cases read, turnaround time, and on-call accuracy,[3,4] but to our knowledge, no studies have analyzed the quantitative trends in fellowship training during an academic year.

We hypothesized that within an academic year, the RTAT for on-call studies dictated by fellows will decrease (ie, improve). Meanwhile, the discrepancy rates will decrease, and clarity of reports will improve. We also hypothesized that participating in independent call will have residual short-term effects on increasing clinical productivity during a subsequent regular work week.

MATERIALS AND METHODS

This study was approved by Hospital of University of Pennsylvania institutional review board.

Fellowship Structure

Each fellowship class consists of approximately 6 first-year fellows. During the 12-month academic year, approximately 7.5 months consisted of fellows interpreting and dictating cross-sectional neuroimaging (CSNI) studies on adult patients from the 3 major hospitals (affiliated with our tertiary care academic institution) during regular workdays, under the supervision of ~18 neuroradiology attendings, with interspersed evening and weekend call. For the remaining 4.5 months, fellows rotated through neurointerventional radiology, pediatric neuroradiology at an affiliated children's hospital, or were off service.

Typical Call Duties

The fellows participated in 3 types of call at our program: weekday neuroradiology evening and overnight call, weekend neuroradiology call, and 1 week of general call.

Primary weekday neuroradiology call duties (Monday–Friday) for fellows at our institution included the interpretation and dictation of emergent and inpatient CT and MR imaging CSNI studies from the 3 major hospitals from 5–11:00 PM. After 11:00 PM, diagnostic radiology residents and overnight radiology fellows dictated all CT studies, while emergent and inpatient MR imaging and CTA studies were dictated by the neuroradiology fellow until 7:00 AM. Most fellows stayed in the hospital until 7:00 AM, though the option was available for fellows to return home and use a "take-home laptop" after 11:00 PM for home pager call (once all work lists for which the fellow was responsible were empty). The take-home laptop contained the same PACS and dictation software as the regular workstations at the hospital and reports were generated in the same manner. Although not explicitly measured, we estimate that >90% of the on-call studies were dictated in-house. Most preliminary reports contained a full dictation, including a clinical history, technique, findings, and impression. One hundred percent of 200 cases randomly audited had full reports. We do not use structured or standardized reporting.

Primary weekend neuroradiology call duties (Saturday and Sunday) for fellows included working with a neuroradiology attending from approximately 7:00 AM to 5:00 PM, during which time the workflow resembled a regular workday (ie, the attending would review studies with the fellow and then finalize the reports with an attestation). From 5:00 PM to 7:00 AM, the fellow would be on independent call and dictate emergent neuroradiology MR imaging and CTA studies in the same manner as in the weekday call.

Coverage of neuroradiology weekday and weekend call was equally split among the fellows throughout the academic year. Each fellow was typically on weekday and/or weekend call once every 1–2 months. An on-call attending neuroradiologist was available as needed but was rarely consulted. Our institution does not use an in-house overnight neuroradiologist.

During the 1 week of overnight general call, the fellow would primarily interpret both neuroradiology and non-neuroradiology CT studies and occasionally plain films and other cross-sectional imaging.

Review of Preliminary Call Reports by Neuroradiology Attendings

Preliminary neuroradiology call reports were reviewed by neuroradiology attendings the following morning and evaluated using a standardized template to assign a degree of discrepancy based on our internally designed system for attending review of preliminary reports:[5]

1. Agree: The attending fully agrees with the report and no words are added.
2. Addition: The attending agrees with the report and adds minor information such as description of a mucosal retention cyst in a paranasal sinus.
3. Minor change: The attending mostly agrees with the report but adds information to clarify the report or correct a mistake, which will not have immediate clinical impact, for example, describing a lesion in the wrong lobe in the cerebral hemisphere or an old orbital blowout fracture.
4. Major change: The fellow missed or misinterpreted a finding that could have immediate clinical impact such as a missed arterial occlusion or acute intracranial hemorrhage.
5. Great call: The fellow detected a subtle finding that could have immediate clinical impact such as detection of a subtle acute infarct on a CT head.

Multiple examinations linked by a single report (eg, CT head and CT cervical spine studies performed concurrently generate 2 separate accession numbers but are dictated as 1 combined report) were assigned the template that captured the highest desired degree of discrepancy (ie, if there was a major miss on the CT head study but not on the CT cervical spine study, the 1 report would be categorized as a "Major change"). All attendings were educated about this scoring system and the templates before finalizing these studies. On average, attendings were in practice 10+ years after completion of their fellowship.

Of note, the attending was not able to edit any portion of the preliminary report dictated by the neuroradiology fellow, including the clinical history, technique, findings, and impression. Additional information and corrections were added by the attending below the preliminary fellow-dictated report.

Call and Noncall Studies

Using our electronic medical record and radiology information system (RIS), we queried all neuroradiology call and noncall

reports generated by neuroradiology fellows on adult patients between July 1, 2012, and June 30, 2017 (60 consecutive months, 29 fellows).

Inclusion criteria for call studies consisted of CSNI reports in which attendings used 1 of the 5 "discrepancy" templates to finalize the study. By contrast, attendings are required to use an "attestation" template during the workday (ie, for noncall studies) to explicitly confirm their participation in the case; such cases were included in the noncall studies pool. Pediatric neuroradiology studies were interpreted at another hospital with a different neuroradiology staff, electronic medical record, and RIS and were excluded. Examinations with RTATs of >1400 minutes (>24 hours) and <0 minutes were excluded because these data entries are nearly uniformly due to errors in "crossover" between various systems (eg, between the dictation software and the RIS).

Determination of RTAT on Call Studies

Our RIS records several workflow-related timestamps for each imaging study in the system. RTAT was defined as the time between verification of images arriving in the PACS to the time of issuance of a full preliminary report by the fellow. RTAT was analyzed separately for CT and MR imaging studies. Of note, the fellow was not responsible for interventional procedures during call shifts. We did not include other factors such as answering phone calls when determining RTAT.

Effects of Taking a Week of Independent Call on Subsequent Clinical Productivity

To measure the effect on the number of studies dictated by a fellow during a regular work week after taking a week of independent call, we analyzed the number of CSNI studies dictated per day by fellows during a regular work week before and after taking a week of independent evening call.

Statistical Analysis

Raw data from the RIS were imported into the R statistical environment for (https://www.r-project.org/) data wrangling and analysis.[6] Each record represented a single imaging study. Available fields in the dataset included provider names (both the fellows and attendings), study (Current Procedural Terminology) codes, multiple procedural timestamps, and the entire radiologic report. Non-neuroradiologic studies were filtered on the basis of the Current Procedural Terminology codes. Minor errors in data entry (eg, misspelling of provider names) were corrected. RTATs were defined as the difference (in minutes) between study completion and the generation of a preliminary report. Using natural language processing, we identified studies performed on call using the text of the discrepancy and attestation templates. These templates also enabled tabulation of the severity of discrepancies between the attending and fellow for call cases (eg, major-versus-minor changes). Finally, we used the position of this attending-inserted template to define the boundary between fellow-generated (above) and attending-generated (below) report text. This feature allowed word and character counts for attending-inserted text for each call case.

Several additional, simple, derived variables were generated on the basis of the raw data, including imaging technique (CT versus MR imaging), day the fellowship study was performed, and the month that a study was completed. Basic statistics were calculated, including total CSNI frequencies, RTATs, and frequencies of report discrepancies. Monthly trends in total call cases, RTATs, and relationships between volume and RTATs were evaluated via regression. We also investigated RTATs for specific imaging studies via general linear regression models that simultaneously accounted for the effects of the day of fellowship, imaging technique (CT versus MR), and interactions. To fully account for the longitudinal nature of the data, we also analyzed RTATs using linear mixed-models treating fellow identification and the day of fellowship as random effects, looking at both linear and nonlinear trends in RTAT with time, ie, allowing each fellow to have a unique nonlinear trajectory in RTAT during the fellow year. Results with both approaches were similar, resulting in identical statistical inferences. Thus, parameter estimates from the more traditional linear regression models are reported here. Multivariable linear regression analysis was used to compare RTAT, word count, and discrepancy rate with the day of fellowship, technique (CT versus MR imaging), daily case load, and RTAT while controlling for individual fellow performance. Similar models were used to compare RTATs between the 11 fellows who completed the diagnostic radiology residency program at our institution versus at other institutions; data from both the first and last 60 days of fellowship were analyzed.

Variables with nonlinear monthly trends were also evaluated with the ANOVA and post hoc Tukey-Kramer tests. Monthly variation in frequencies of call-discrepancy templates were assessed via χ^2 tests. Changes in frequencies of specific template use between fellowship semesters were assessed via serial 2-sample tests of proportions.

Finally, we attempted to determine whether the presence of a week of call influences case volumes during subsequent day shifts. We constructed a crude measure comparing the total number of cases for the week before and after each week-long call shift. Given their non-normal distribution, differences in the number of cases read before and after call were assessed via the Wilcoxon test. For testing our hypotheses, statistical significance was defined as $P < .05$.

RESULTS

Durng 5 academic years, data from 12,072 CSNI call studies dictated by fellows were analyzed. One hundred forty-six studies (1.2%) were excluded due to 0 minutes<RTAT >1400 minutes. Twenty-nine fellows dictated 11,926 CSNI call studies with a mean annual call volume of 411.2 [SD, 306] per fellow. Fellows participated in 53 median annual call shifts that covered adult CSNI: mean total call studies = 18.29 [SD, 8.12]; mean MR imaging call studies = 7.03 [SD. 4.14]; and mean CT call studies = 10.28 [SD, 6.39].

Call Cases per Month

Trends in call volumes per month are summarized in Fig 1. The number of call cases dictated per month increased significantly throughout the academic year ($F^1_{10} = 37.61$, P value = .001, reaching a maximum of 1448 total cases in June (48.3 cases/day; MR imaging = 504 [16.8 cases/day] and CT = 944 [31.5 cases/day]). When they were measured separately, there were

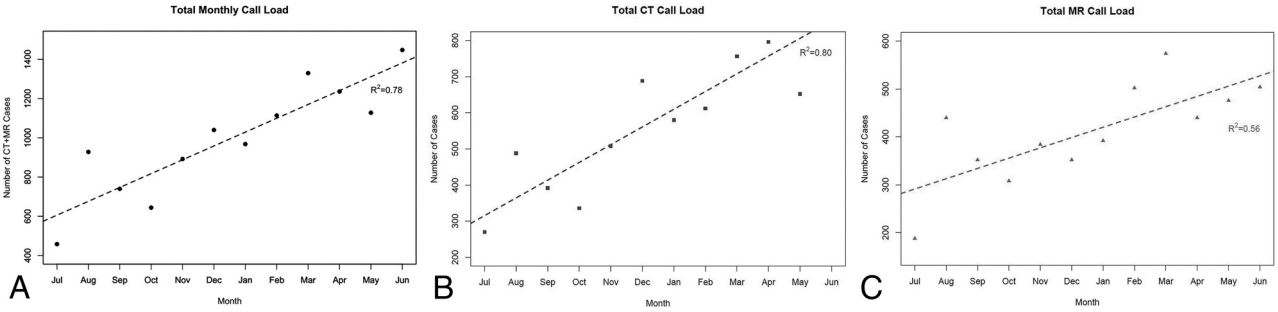

FIG 1. *A*, Mean total (CT + MR imaging) call cases per month aggregated over 5 fellowship classes with R^2 values. *B*, Mean total (CT) call cases per month aggregated over 5 fellowship classes with R^2 values. *C*, Mean total (MR imaging) call cases per month aggregated over 5 fellowship classes with R^2 values.

statistically significant increases in both monthly MR imaging ($F^1_{10} = 12.71$, *P* value = .005) and CT volumes ($F^1_{10} = 38.81$, *P* < .0001). The rate of CT studies dictated during the year increased significantly faster than that of MR imaging ($F = 7.734$, *P* value = .011). Fellows interpreted a mean of 72 call MR imaging studies during their first 6 months [SD, 60] and 100 MR imaging studies [SD, 70] during the last 6 months of fellowship, and a mean of 96 call CT studies during the first 6 months [SD, 102] and 150 studies [SD, 121] during their final 6 months.

Report Turnaround Times

Most preliminary reports were generated by fellows within 3 hours of study completion (RTAT mean = 136 minutes; 95% CI, 132–139 minutes; median = 72 minutes; 95% CI, 70–74 minutes), though the mean was skewed upward by (relatively infrequent) studies with very long RTATs. In general, MR imaging studies required significantly more time to complete than CT ($\chi^2 = 901.0$, $df = 1$, *P* value = .0001); during the study interval, the mean time to generate a preliminary MR imaging report was 183 minutes (95% CI, 177.1–188.8 minutes; median, 111 minutes; 95% CI, 107–116 minutes), while for CT, the mean RTAT was 104 minutes (95% CI, 100.9–107.1 minutes; median, 55 minutes; 95% CI, 54–57 minutes). Days since beginning fellowship were significantly associated with faster mean RTAT ($F^1_{11924} = 136.2$, *P* value < .0001; $\beta = -0.19$, *P* value < .0001), with each day of fellowship estimated to improve RTATs by 0.18 minutes in our regression models. Hierarchic linear mixed-models (Fig 2*D*) found significant variability of RTAT trajectories among fellows during the academic year, with some fellows demonstrating minimal improvement or even increases with time. However, most fellows improved with increased training, as did the group mean. There was a statistically significant interaction between imaging technique and day of fellowship ($\beta = -0.07$, $t = -2.39$, $P = .01$), with RTATs for MR imaging decreasing faster than for CT (−0.15 min/day for CT and −0.22 min/day for MR imaging). RTAT was 67.7 minutes slower for MR imaging compared with CT (*P* value ≤ .0001).

When data were aggregated by month, there were small-but-statistically significant decreases in both mean ($F^1_{10} = 7.25$, *P* value = .023) and median ($F^1_{10} = 12.16$, *P* value = .006) monthly RTATs during the fellowship year (Fig 2*A*). The pattern was not entirely monotonic, however, with the RTAT nadir seen in mid-fellowship around December (median RTAT = 58 minutes). When month was treated as a categoric rather than continuous variable, ANOVA suggested significant monthly variation ($F^{11}_{11816} = 34.04$, *P* value < .0001). Post hoc tests found that RTATs were significantly faster in December relative to all other months, with the exception of May and June. From July (first month of fellowship) to June (12th and last month of fellowship), the mean report turnaround time for all CSNI studies decreased from 144.4 to 99.4 minutes (−45 minutes, −31%) and median report turnaround time decreased from 79 to 55 minutes (−24 minutes, −30%).

We hypothesized that average RTAT would be associated with call volumes, with busier months associated with slower rates of reporting. However, we observed that call volumes were inversely related to RTATs (Fig 2*C*). Specifically, RTATs for CT studies significantly decreased with increasing case volumes ($r = -0.70$, $F^1_{10} = 9.639$, *P* value = .011), while the RTATs for MR imaging cases decreased slightly but were not statistically significant ($r = -0.13$, F^1_{10}, 0.1782, *P* value = .682). Linear regression demonstrated that RTATs decreased by 0.44 minutes for each extra case dictated that day (*P* value = .004).

Finally, we compared the RTAT for fellows who completed a diagnostic radiology residency at our institution ($n = 11$) with fellows who completed it at another institution ($n = 18$). Linear regression analysis demonstrated that fellows who completed a radiology residency at our institution had significantly lower RTATs during the first 60 days of fellowship (*P* value ≤ .0001); however, the effect persisted during the year, with a similar group difference in the last 60 days of fellowship (*P* value < .0001).

Discrepancies in CSNI Interpretation between Fellows and Attendings

The monthly proportion of "Great call," "Agree," "Addition," "Minor change," and "Major change" template use varied significantly during the academic year ($\chi^2 = 517.07$, $df = 44$, *P* value < .0001). In general, the frequencies of less discrepant codes increased during the year, with corresponding decreases in templates associated with more substantial discrepancies. The lowest percentages of Major and Minor changes were seen during the seventh month of fellowship (January, 0% and 4.5%, respectively). When data were aggregated by semester, significant changes were seen between the first 6 months of training compared with the last 6 months for most categories (Tables 1 and 2). Multivariable linear regression demonstrated that the call discrepancies slightly decreased with each day of fellowship (*P* ≤

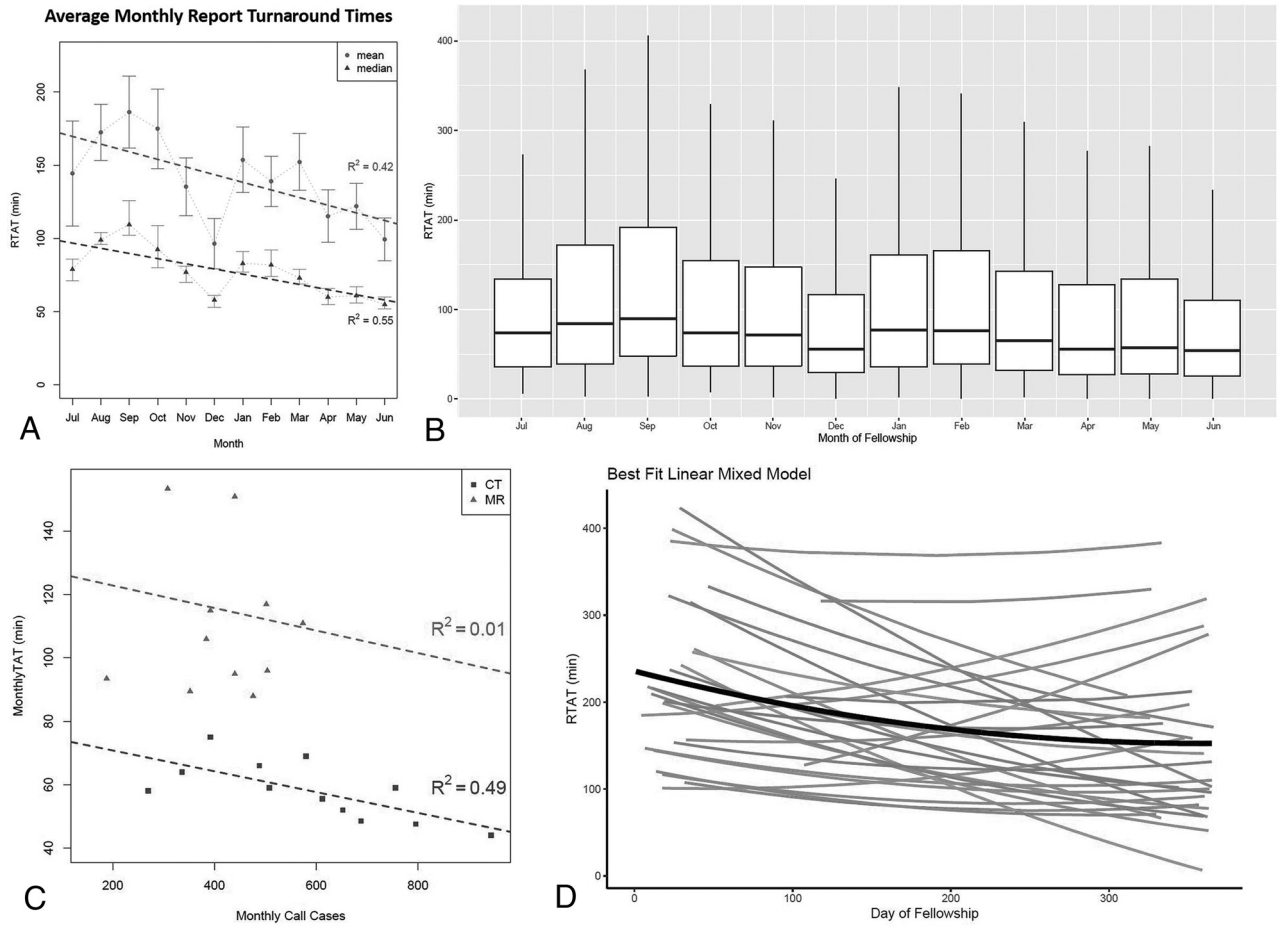

FIG 2. *A*, Average monthly turnaround times for CT and MR studies during fellows' call. Mean (*circle*) and median (*triangle*) values are both shown. *Dashed lines* represent best-fit regression lines during the academic year, while month-to-month changes are shown by the *dotted lines*. R^2values are provided. *B*, Box-and-whisker plot demonstrates the RTAT for each month. *C*, Relationships between monthly call volume and monthly median RTATs for CT (*square*) and MRI (*triangle*) separately. R^2 values are provided. *D*, *Lines* depict individual trajectories in RTAT during the fellowship year for each fellow based on a linear mixed-model. The *bold line* shows the curve for the aggregated data.

Table 1: Median RTATs by quarters of the academic year

Quarter	RTAT (Min)
1st (July-September)	97 (95% CI, 91–102)
2nd (October-December)	68 (95% CI, 65–73)
3rd (January-March)	79 (95% CI, 75–83)
4th (April-June)	59 (95% CI, 56–62)

Table 2: Median RTAT and CI for December and January

Month	RTAT (Min)
December	58 (95% CI, 53–61)
January	83 (95% CI, 77–91)

.001). With an increasing daily case load, the number of discrepancies slightly increased ($P \leq .0001$). Technique (eg, CT versus MR imaging) and RTAT had no statistically significant effect on call discrepancies. We did an ordered logistic regression for graduation years 2015–2017 and looked at the day of fellowship, imaging technique (CT versus MR imaging), and daily case load as predictors of the severity of call discrepancy. There was a significant effect of graduation year (logistic regression, 39; *P* value $< .0001$). We looked at the discrepancy data for each year and saw a generalized decline in the proportion of Minor changes, an increase in Agree + Addition, and no significant change in the proportion of Major changes or Great calls. We looked at 131 major misses, 100% of the total major misses, 53% CT and 47% MR imaging. Most (75%) were perception errors, while 25% were misinterpretation errors and overcalls. Of the misses, there were 14% missed thrombus/occlusion, 9% missed infarcts, 9% missed intracranial hemorrhages, 6% missed fractures, and 6% missed aneurysms. The remaining 56% of cases had a variety of perceptually missed or misinterpreted findings.

Quantitation of Attending Additions to Final Reports

On average, neuroradiology attendings added 141 words (95% CI, 137–145) and 1015 (95% CI, 987–1044) characters to reports coded "Addition," 184 (95% CI, 173–195) words and 1305 (95% CI, 1227–1383) characters to those with Minor changes and 188 (95% CI, 165–211) words and 1304 (95% CI, 1142–1469) characters to those with Major changes. There were significant changes in monthly rates of both words ($F^{11}_{12060} = 8.05$, *P* value $< .001$) and characters ($F^{11}_{12060} = 9.44$, *P* value $< .001$) addended to preliminary reports, with the lowest number of additions seen during the sixth

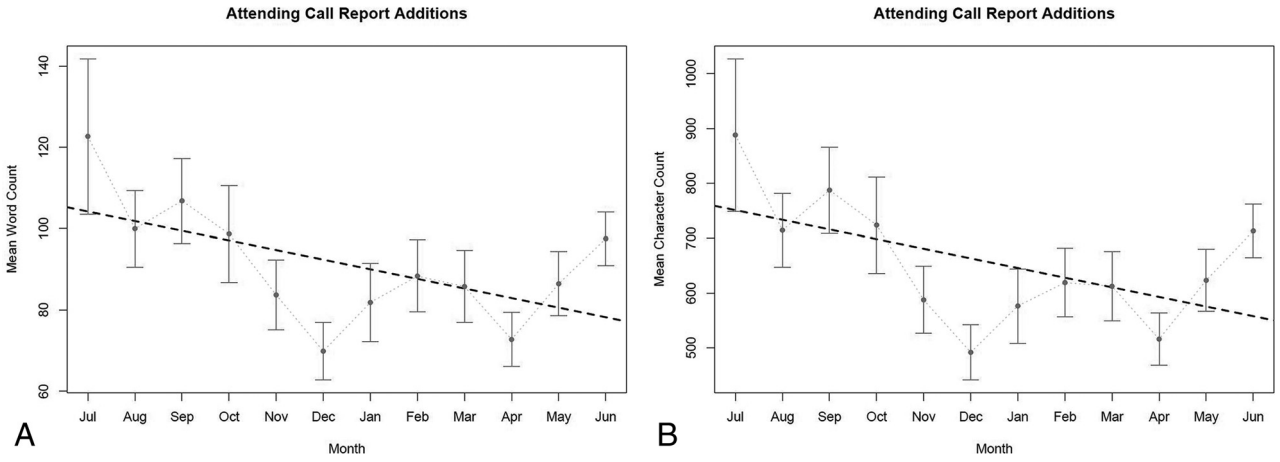

FIG 3. *A*, Mean word count added to call reports per month of fellowship of training. Confidence intervals are shown with *bars*. *B*, The mean character count added to call reports per month of fellowship training. Confidence intervals are shown with *bars*.

month (December; Fig 3). When data were aggregated into semesters, the second 6 months of the fellowship year had significantly fewer mean additions of both words, 93.5 (95% CI, 89–98) versus 86.2 (95% CI, 83–89; $F^{1}_{12070} = 7.28$, P value = .007) and characters, 672 (95% CI, 641–702) versus 616 (95% CI, 593–640; $F^{11}_{12060} = 9.44$, P value < .001) compared with the first 6 months of fellowship. Linear regression demonstrated that for each day of fellowship, the number of words added to preliminary reports decreased by 0.154 (P value ≤ .0001). The number of words added to MR imaging reports was greater by 39.4 compared with CT reports (P value ≤ .0001). Daily case load and RTAT did not significantly affect the number of words added.

Effect of Taking Week-Long Call on the Number of Studies Dictated during the Subsequent Week

There was a slight but not statistically significant decrease in number of studies dictated per day (8 studies/day) a week (7 days) after taking a week of call (5 days) compared with the prior week (9 studies/day for 7 days) (Wilcoxon test W = 1295598, P value = .06).

DISCUSSION

The results of this study demonstrate that progress through a 1-year neuroradiology fellowship at a large academic institution with participation in independent call correlates with the following: 1) a decrease in RTAT for independent call studies, 2) a decrease in discrepant findings and an increase in agreement with the attendings' final interpretation of independent call studies, and 3) a decrease in words added to correct or clarify descriptions of pertinent findings in independent call reports despite an increase in call volume throughout the academic year. While we are aware that neuroradiology fellowship programs are variable in their education and call structures, we believe that all programs have some type of call component as well as attendings with varying reporting styles and levels of interest in fellows' education; thus, our findings are generalizable to objectively monitoring fellowship progress.

Reducing RTAT for radiology studies has become an essential goal for both academic and nonacademic radiology practices in the past 2 decades.[7] Studies have analyzed RTATs for radiology residents;[3,8] however, our findings are unique because they investigate RTAT during 1 academic year in a postresidency fellowship. Most fellows are employed radiologists in radiology practices on graduation, and their duties often center on interpretation of CSNI studies,[9] with reimbursement and advancement potentially tied to RTATs.[7] Hence, it is important that knowledge of trends in RTAT for call studies be available to fellows and hiring radiology practices. Our findings demonstrate a gradual decline in RTATs for both MR imaging (−0.22 min/day) and CT (−0.15 min/day) reports during the academic year, with a nadir in December (midpoint of the academic year, median RTAT of 58 minutes), and a 30% decrease in RTAT for all CSNI studies from the first (July) to the last month (June) of the academic year, indicative of growing proficiency in interpreting CSNI studies. Our results are consistent with a prior study that demonstrated decreased RTATs after changing from general to subspecialty reporting,[10] noting that the decrease in our study was nonlinear, as shown in Fig 2*B* and -*D*. The RTATs for CSNI MR imaging studies remained higher than for CTs throughout the academic year, a finding we attribute to the increased number of images per study and their overall complexity. This finding is congruent with a prior study that demonstrated that neuroradiology fellows interpret and dictate MR imaging brain reports in an average of 18 minutes, which is significantly higher than for CTs.[11] The RTAT for fellows who completed diagnostic radiology residency at our institution remained significantly lower for the first 60 and last 60 days of the fellowship compared to fellows who did not complete a diagnostic radiology residency at our institution. The reason is probably multifactorial and related to familiarity with the dictation, PACS, and hospital electronic medical record systems and the attendings.

We hypothesized that increasing volumes on call would cause an increase in RTAT. However, our findings demonstrate an inverse relationship, with a statistically significant decline in RTATs for CT studies and a trend toward a decline for MRIs. Radiology residents are typically exposed to higher numbers of CT studies during residency compared with MRIs; therefore, new fellows start with a relative proficiency in this technique. This

Table 3: Semiannual change in proportions of call discrepancies between fellows and attendings

Template	July-December (p_A)	January-June (p_B)	Difference	Hypothesis Tests (H_0: $p_A = p_B$)
Great call	0.7%	0.9%	+0.2%	$\chi^2 = 1.26$, P value $< .262$
Agree	33.4%	42.8%	+9.4%	$\chi^2 = 106.17$, P value $< .0001$
Addition	52.2%	46.7%	–5.4%	$\chi^2 = 33.57$, P value $< .0001$
Minor change	12.3%	8.2%	–3.7%	$\chi^2 = 42.27$, P value $< .0001$
Major change	1.4%	0.9%	–0.5%	$\chi^2 = 6.61$, P value = .010

Note:—p_A indicates proportions of call discrepancy during the first half of the academic year; p_B, proportions of call discrepancy during the second half of the academic year; H_0, null hypothesis.

"experience differential" between modalities may be particularly pertinent in the call setting because many residents have minimal experience in interpreting MR imaging studies independently, and MR imaging studies typically contain more images and are often more complex. On high-volume call nights, fellows may more easily adapt to expanding CT worklists due to a greater experience with this technique, while their relative inexperience in MR imaging attenuates the ability to reduce RTAT with increasing case volumes. We also believe that RTAT decreased as case volumes increased due to necessity: If not busy, then fellows can spend more time reviewing and dictating cases. Of note, our fellowship program does not actively track or emphasize RTAT; therefore, fellows did not have pressure to rapidly interpret studies and generate reports on the basis of productivity benchmarks. Anecdotally, some fellowship programs track RTAT, and our data can help these programs assess the progress of their fellows by following the data shown in Tables 1 and 2 and Fig 2, noting that large variabilities in RTAT trends can be seen among fellows. Care must be taken because overemphasis on RTAT can be detrimental and might encourage the fellow to work out of their comfort zone to quickly interpret and generate a report, potentially leading to more interpretive errors. A prior study demonstrated that required RTAT negatively affected a radiologist's ability to teach and the quality of resident education.[12]

In many radiology residency and fellowship programs, radiology trainee performance on call is objectively evaluated with discrepancy rates. The major discrepancy rate for our fellows was 1.1%, which is overall lower than for residents[3,8] and lower than rates for neuroradiology fellows reported by another institution (4.8%).[13] Our findings demonstrate decreases in major and minor changes during the academic year per day, with the nadir during January (around the midpoint of the fellowship year). To our knowledge, no prior study has analyzed the discrepancy rate trends in fellows, but declines in discrepancy rates were reported in a prior study for radiology residents assessed for 5 years.[4] Our findings also demonstrate that the percentage of the Addition template usage decreased and the percentage of Agree template usage increased, all while the monthly call load increased. This finding suggests that fellows' skills overall improved in identifying the pertinent findings and making accurate diagnoses during the course the academic year so that attendings did not need to add any extra wording to the report. Overall usage of the Great call template was low but increased slightly during the last 6 months of training. A study published in 2017 demonstrated that 25%, 62%, and 13% of surveyed neuroradiologists who trained fellows in the United States believed that fellows were less capable, equally capable, or more capable, respectively, compared with prior years.[2] We analyzed data from 2015 to 2017 (ie, close to the time of the survey) and saw a generalized slight decline in the proportion of Minor changes, an increase in Agree and Addition templates, and no change in Major change and Great call templates, suggesting no increase in discrepancy rates in our fellows. Caution must be used when comparing our findings with those in the prior survey because we looked only at discrepancy rates at 1 institution, while the survey asked a global question about overall capabilities of fellows.

The data in Table 3 can be used by fellowship programs to assess progress at the end of December, at which time, on average, during 6 months, approximately 86% of reports would have agreement or the addition of relatively minor information, approximately 12% of reports would have minor changes, and approximately 1% of studies would have major changes. The percentages in each category can be combined as needed because some programs might not differentiate between major or minor discrepancies or between reports with words added by the attending and general agreement.

We sought to determine whether other variables associated with call could affect the discrepancy rate. The discrepancy rate slightly increased with an increasing daily case work load (ie, more cases on call results in a higher discrepancy rate) and is congruent with a prior study that demonstrated that higher daily case volumes are associated with higher discrepancy rates for neuroradiologists.[14] Fatigue from a busy clinical day is a likely contributory factor as well. RTAT and imaging technique (CT versus MR imaging) had no effect on the discrepancy rate. We analyzed 100% of the reports with a major discrepancy and found that 75% of cases had a perception error and 25% had a misinterpretation or overcall error. Most interesting, our findings are very similar to those in a prior study that demonstrated that for discrepant neuroradiology attending reports, 74.8% were perception errors and 25.2% interpretation errors.[15]

The quality and clarity of neuroradiology reports are important factors that clinicians use to judge the value a neuroradiologist provides in clinical practice.[15] Our study indirectly evaluated the clarity of preliminary call study reports quantitatively by determining the number of words and characters added to the end of reports. Our neuroradiology attendings regularly edit fellow reports during regular workdays to correct or emphasize relevant findings and/or for style. However, for call studies, attendings cannot change, add, or delete words in the preliminary report and are instructed and expected to only add words to the end of the fellow's report to clarify, change, or emphasize pertinent findings regardless of the report style or the fellow's level of training. Therefore, we believe that the number of words and characters added to call reports is a surrogate marker for the

clarity of a report. Because fellows generally did not review the findings with an attending before dictation, a study with no or minimal addition of words reflects an independently well-constructed, accurate, and coherent report, an objective metric for progression in fellowship. Our method is a variation of analysis of percentage change characters used by a prior study[16] to measure the quality of reports by trainees and has been shown to be a meaningful measure of trainee progress in residency.[17] On these prior studies, a distinction between when trainees had discussed the findings with the attending before generating a report compared with a solo interpretation and report generation was not always made. In contrast, in our study, for nearly all studies, the fellow interpreted the study and generated a report without prior input from an attending.

Our study demonstrates that ~29% more characters were added to reports with major and minor changes (ie, missed important findings) compared with reports with the Addition templates, while an earlier study demonstrated a 41.6% character change for studies with missed critical findings.[16] Although we cannot directly compare the data, we find it interesting that both studies demonstrated an increase in words and characters used to denote a missed finding. Studies with critical misses often have complex and multiple findings requiring more verbiage to effectively describe them. The number of words and characters attendings added at the end of the preliminary call reports reached a nadir during December and decreased by approximately 8% in the second semester, despite an overall increase in call volume and a decrease in RTAT. From July to December, a mean of approximately 94 words and 672 characters was added to reports. These data could be used by programs to assess whether fellows are within a range of the norm at the end of December or might require additional training to increase the quality and clarity of their reports.

The etiology of this nadir in December is likely multifactorial and probably related to holiday workflow (though the number of studies dictated during December and January did not decline) as well as the fellow's increasing skill level, comfort level with attendings, and confidence in interpreting studies and generating reports. During the second half of the fellowship, many fellows have already procured a postfellowship job, which, therefore, possibly reduced the motivation to impress the attendings (ie, generating the "perfect" call report) and could be a contributing factor. We do not believe attending clinical duties would impact the number of words added because each regular work day has a dedicated attending assigned to finalize fellow preliminary reports without pressure of interpreting mounting cases on that day (ie, there is no real incentive to minimize time spent on each case).

We sought to determine whether other variables associated with call might affect the number of words added to the preliminary report. MR imaging reports had significantly more words added compared with CT; and this feature is likely due to the more complex information obtained from MR imaging compared with CT, requiring longer and often more detailed reports. Most interesting, the daily case load and RTAT had no effect on the number of words added.

Finally, we wanted to determine whether participating in weekly independent call had any residual effect on the number of studies dictated during a regular subsequent work week. We hypothesized that due to the autonomy practiced during call, the fellows would become more efficient in interpreting and dictating CSNI studies and this outcome would lead to higher clinical productivity, eg, more cases dictated. However, our data demonstrate no significant difference in the number of cases dictated during a regular work week before and after taking a week of call. The reason for this finding is probably multifactorial but is likely, in part, because fellows function differently during the regular workday (honing skills) rather than on call (applying skills). Postcall fatigue may also be a contributory factor.

Overall increases in CSNI call volumes seen at our institution fit with recent national trends.[18] The monthly increase in call volume is probably multifactorial and attributable to inclusion of new criteria for stroke imaging and the addition of overnight MR imaging technologists, among others. Analysis of these factors is beyond the scope of our article.

Our study has limitations that must be considered when interpreting our results. First, this study uses a retrospective cohort design, which could be affected by confounding factors including incomplete retrieval of data due to technical factors. Given the large sample of >10,000 cases, we could not easily assess data integrity for all data points. Additionally, we believe the possible exclusion of some studies would likely not substantially affect the trends shown in our article because we could not identify a systematic bias in patterns of missing data. Second, we did not differentiate the type of CT and MR imaging studies or discrepancies on the basis of technique; however, the complexity of cases can be highly variable irrespective of technique. Third, we did not determine the reasons for the discrepancies in call reports (ie, perceptual-versus-cognitive errors) or the clinical impact of the words added by the attendings at the end of the preliminary report. Fourth, we did not look at RTAT, discrepancy rates, or words added for noncall reports. Although some of these data from another institution were addressed in a prior study.[17] Fifth, there are no criterion standards with which to compare our data with regard to ideal RTAT, discrepancy rate, or words added to reports. Sixth, we could not correlate the discrepancy rates or number of words added to the report with individual attendings. Due to technical factors related to storing preliminary fellow reports in our RIS software, 89% of call reports did not have the name of the attending who finalized the report. However, given that our section had a relatively stable cohort of attendings, their clinical duties were relatively equally distributed, and they were instructed specifically on when and how to addend reports, we believe this variability is largely diminished.

CONCLUSIONS

Progression through neuroradiology fellowship and independent call correlates with decreased discrepancies and improvement in indirect quantitative measurements of the clarity of call reports, significant decreases in RTAT for CT studies, and some decrease in RTAT for MR imaging studies, despite increases in call volumes during the academic year. These metrics can be tracked throughout the academic year, and the midyear would be a

logical time point for programs to assess objective progress of fellows and address any deficiencies.

Disclosures: J. Eric Schmitt—*UNRELATED*: *Travel/Accommodations/Meeting Expenses Unrelated to Activities Listed*: Radiological Society of North America, *Comments*: waived registration for teaching a statistical computing course. Tessa S. Cook—*UNRELATED*: *Board Membership*: Society for Imaging Informatics in Medicine, Association of University Radiologists, Pennsylvania Radiological Society, Philadelphia Roentgen Ray Society, *Comments*: travel reimbursement for meetings; *Grants/Grants Pending*: Radiological Society of North America, American College of Radiology, Society for Imaging Informatics in Medicine, American College of Radiology Imaging Network*; *Royalties*: Osler Institute, *Comments*: royalties for lectures originally given and recorded in 2012 on cardiac imaging. Suyash Mohan—*UNRELATED*: *Consultancy*: Northwest Biotherapeutics; *Grants/Grants Pending*: Novocure, Galileo.* *Money paid to the institution.

REFERENCES

1. **ACGME Program Requirements for Graduate Medical Education in Neuroradiology.** July 1, 2020. https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/423_Neuroradiology_2020.pdf?ver=2020-06-29-164131-690. Accessed July 15, 2020
2. Chen JY, Lexa FJ. **Baseline survey of the neuroradiology work environment in the United States with reported trends in clinical work, nonclinical work, perceptions of trainees, and burnout metrics.** *AJNR Am J Neuroradiol* 2017;38:1284–91 CrossRef Medline
3. Shah NA, Hoch M, Willis A, et al. **Correlation among on-call resident study volume, discrepancy rate, and turnaround time.** *Acad Radiol* 2010;17:1190–94 CrossRef Medline
4. Sistrom C, Deitte L. **Factors affecting attending agreement with resident early readings of computed tomography and magnetic resonance imaging of the head, neck, and spine.** *Acad Radiology* 2008;15:934–41 CrossRef Medline
5. Chen PH, Chen YJ, Cook TS. **Capricorn: a Web-based automatic case log and volume analytics for diagnostic radiology residents.** *Acad Radiol* 2015;22:1242–51 CrossRef Medline
6. Core Team R. **R: A language and environment for statistical computing.** R Foundation for Statistical. Computing. https://www.R-project.org/. Accessed August 1, 2020
7. Giles WL, Boland EF, Halpern G. **Radiologist report turnaround time: impact of pay-for-performance measures.** *AJR Am J Roentgenol* 2010;195:707–11 CrossRef
8. Ruutiainen AT, Scanlon MH, Itri JN. **Identifying benchmarks for discrepancy rates in preliminary interpretations provided by radiology trainees at an academic institution.** *J Am Coll Radiol* 2011;8:644–48 CrossRef Medline
9. Rosenkrantz AB, Wang W, Hughes DR, et al. **Generalist versus subspecialist characteristics of the U.S. radiologist workforce.** *Radiology* 2018;286:929–37 CrossRef Medline
10. Stern C, Boehm T, Burkhardt S, et al. **Subspecialized radiological reporting expedites turnaround time of radiology reports and increases productivity.** *Rofo* 2018;190:623–29 CrossRef Medline
11. Al Yassin A, Sadaghiani MS, Mohan S, et al. **It is about "time": academic neuroradiologist time distribution for interpreting brain MRIs.** *Acad Radiol* 2018;25:1521–25 CrossRef Medline
12. England E, Collins J, White RD, et al. **Radiology report turnaround time: effect on resident education.** *Acad Radiol* 2015;22:662–67 CrossRef Medline
13. 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
14. Patel SH, Stanton CL, Miller SG, et al. **Risk factors for perceptual-versus-interpretative errors in diagnostic neuroradiology.** *AJNR Am J Neuroradiol* 2019;40:1252–56 CrossRef Medline
15. Olthof AW, De Groot JC, Zorgdrager AN, et al. **Perception of radiology reporting efficacy by neurologists in general and university hospitals.** *Clin Radiol* 2018;73:675.e1–e7 CrossRef Medline
16. Stankiewicz K, Cohen M, Carone M, et al. **Comparing preliminary and final neuroradiology reports: what factors determine the differences?** *AJNR Am J Neuroradiol* 2016;37:1977–82 CrossRef Medline
17. 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
18. Verdoorn JT, Hunt CH, Luetmer MT, et al. **Increasing neuroradiology exam volumes on-call do not result in increased major discrepancies in primary reads performed by residents.** *Open Neuroimag J* 2014;8:11–15 CrossRef Medline