QRSE: a Novel Metric for the Evaluation of Trainee Radiologist Reporting Skills

David Surrey · Richard E. Sharpe Jr. ·
Richard J. T. Gorniak · Levon N. Nazarian ·
Vijay M. Rao · Adam E. Flanders

Published online: 5 February 2013
© Society for Imaging Informatics in Medicine 2013

Abstract Diagnostic radiology training programs must produce highly skilled diagnostic radiologists capable of interpreting radiological examinations and communicating results to clinicians. Established training performance tools evaluate interpretive skills, but trainees' competency in reporting skills is also essential. Our semi-automated passive electronic tool entitled the Quantitative Reporting Skills Evaluation (QRSE) allows radiology training programs to evaluate the quantity of edits made to trainee preliminary reports by attending physicians as a metric to evaluate trainee reporting performance. Consecutive report pairs and metadata extracted from the radiology information system were anonymized and exported to a MySQL database. To perform the QRSE, for each report pair, open source software was first utilized to calculate the Levenshtein Percent (LP), the percent of character changes required to convert each preliminary report to its corresponding final report. The average LP (ALP), ALP for each trainee, and standard deviations were calculated. Eighty-four trainees and 56 attending radiologists interpreted 228,543 radiological examinations during the study period. The overall ALP was 6.38 %. Trainee-specific ALPs ranged from 1.1 to 15.3 %. Among trainee-specific ALPs, the standard deviation was 3.7 %. Our analysis identified five trainees with trainee-specific ALPs above 2 standard deviations from the mean and 14 trainees with trainee-specific ALPs less than 1 standard deviation below the mean. The QRSE methodology allows for the passive, quantitative, and longitudinal

evaluation of the reporting skills of trainees during diagnostic radiology residency training. The QRSE identifies trainees with high and low levels of edits to their preliminary reports, as a marker for trainee overall reporting skills, and thus represents a novel performance metric for radiology training programs.

Keywords Reporting · Internship and residency · Medical Education · Electronic medical records · Open source · Radiology reporting · Radiology teaching files · Radiology information systems

Introduction

Diagnostic radiology training programs strive to produce highly skilled diagnostic radiologists capable of interpreting radiological examinations and communicating results to clinicians. The radiology report is the end product of the radiologist's interpretation and often the only contact the radiologist has with the referring provider. There is an extensive body of literature describing the expected contents of the radiology report [1–4].

Conventional educational evaluation tools, such as the American College of Radiology Diagnostic Radiology In-Training exam [5] and the American Board of Radiology Board Initial Certification Examinations [6] focus primarily on evaluating a trainee's interpretive skills and radiological fund of knowledge rather than evaluating his/her ability to produce a succinct and coherent radiology report. Consequently, we sought to create a quantitative metric to passively and quantitatively evaluate trainees' reporting skills.

Within radiology training programs, individual attending radiologists educate and mentor trainee radiologists in the art of interpreting radiological examinations and in drafting accurate and succinct imaging reports. As part of the

This manuscript was presented at the 2012 SIIM Annual Meeting and received the Roger A. Bauman, MD Award.

D. Surrey · R. E. Sharpe Jr. (✉) · R. J. T. Gorniak ·
L. N. Nazarian · V. M. Rao · A. E. Flanders
Department of Radiology, Thomas Jefferson University Hospital,
132 South 10th Street, Main Building 10th Floor,
Philadelphia 19107 PA, USA
e-mail: richsharpejr@gmail.com

standard workflow in radiology training programs, radiology trainees draft preliminary reports that are subsequently reviewed, and may be edited, by attending radiologists prior to report finalization.

During a typical residency experience, trainees submit thousands of preliminary reports to various attending radiologists within a radiology department. For each preliminary report submitted by each trainee radiologist, each attending radiologist independently reviews and evaluates the need for report modifications. Attending radiologists may make more edits to some trainees' reports than to other trainees' reports, and these differences may reflect varying levels of trainee reporting skills. For example, trainees whose reports are rarely edited may possess higher levels of reporting skills, and trainees whose reports are consistently heavily edited by attending radiologists may need further mentoring and education in reporting skills. To our knowledge, there exists no methodology for longitudinal quantitative evaluation of reporting skills of diagnostic radiology trainees.

We created a semi-automated passive electronic educational tool, the Quantitative Reporting Skills Evaluation (QRSE), which allows radiology training programs to evaluate the quantity of edits made to trainee preliminary reports by attending physicians as a metric to evaluate trainee reporting performance.

Materials and Methods

Institutional review board approval was first obtained prior to this investigation. All preliminary and finalized radiology report pairs from examinations performed from March 14, 2010, to July 13, 2010, and from December 31, 2010, to February 12, 2011, were extracted from the radiology information system (RIS, GE Centricity) and exported to a database (MySQL). A vendor-applied RIS patch precluded data collection from July 14, 2010, to December 30, 2010. Anonymized attending and trainee radiologist identifiers, as well as report metadata, such as report times, imaging modality, body section, and radiological subspecialty, were also extracted.

To perform the QRSE, open source software [7] was first utilized to calculate the Levenshtein Distance (LD), for each report pair. The LD is a numerical result indicating the number of character changes required to convert each preliminary report to its corresponding final report. Subsequently, the report length was calculated for each preliminary and finalized report pair. The Levenshtein Percent (LP) was calculated for each report by dividing the corresponding LD by the longer of either the preliminary or final report total character length. The LP normalizes for differences in the overall length of each report and represents the percentage of characters changed

in a given report to convert a preliminary report to its corresponding finalized report.

LP values for all included reports were analyzed using SPSS and Microsoft Excel. The average LP (ALP), standard deviation, 95 % confidence intervals, minimum, median, and maximums were calculated for trainees overall. The number of examinations interpreted by trainee radiologists was also analyzed and plotted alongside each trainee-specific ALP.

Results

During the study period, 84 trainees and 56 attending radiologists interpreted 228,543 radiological examinations. Each trainee reported an average of 2,720.8 examinations during the study period. The maximum number of cases interpreted by any single trainee during the study period was 7,474 and the minimum was 513.

For all examinations, the ALP was 6.38 % with a 95 % confidence interval of 6.31–6.43 %. A total of 128,536 (56.2 %) reports were not changed during the process of report finalization. Figure 1 is a histogram demonstrating the extent of edits made to reports that were changed during the report finalization process.

Analysis of trainee-specific ALPs demonstrated a mean ALP of 6.4 %, 95 % CI of 5.6–7.2 %, minimum ALP of 1.1 %, median ALP of 5.9 %, and maximum ALP of 15.3 %. The standard deviation of trainee-specific ALPs was 3.7 %. The ratio of the highest trainee average LP over the lowest trainee average LP was 13.9 (15.3 %/1.1 %). Figure 2 demonstrates the number of examinations interpreted by each trainee, as well as each trainee's ALP.

Fig. 1 Histogram of quantity of edits made to reports changed during the report finalization process

Fig. 2 ALPs (with 95 % CI) for each trainee are plotted alongside the number of examinations reported by each trainee. Trainee ALP, the average percent of characters changed during report finalization,

ranged from 1.1 to 15.3 %, with a mean of 6.4 %. Trainees interpreted a minimum of 513 examinations, maximum of 7,474 examinations, and an average of 2,720.8 examinations during the study period

QRSE analysis identified 5 trainees (5.9 % of trainees) with trainee-specific ALPs above 2 standard deviations from the mean and 12 trainees (14.3 % of trainees) with trainee-specific ALPs between 1 and 2 standard deviations above the mean. Fourteen trainees (16.7 % of trainees) demonstrated trainee-specific ALPs less than 1 standard deviation below the mean.

Discussion

Attending physicians of diagnostic radiology training programs are tasked with producing residents who are competent to serve as diagnostic radiologists. Several conventional metrics assist program directors in accomplishing this task, such as faculty evaluations and peer evaluations, as well as scores from ACR in service and ABR certification examinations.

While a few additional tools have recently been proposed to assist in the evaluation of trainee performance, this manuscript presents the only known longitudinal quantitative metric to analyze the extent of changes made to trainee preliminary reports by attending radiologists.

Several studies have evaluated discrepancies from the interpretations of on-call examinations interpreted independently by trainees, and these studies have established baseline expected discrepancy rates between attending and trainee radiologist interpretations [8–11]. Recently, Itri et al. described a web-based application (Orion) to quantify the discrepancy rates among a subset of trainee preliminary reports [12]. Orion allows for evaluation of individual and collective trainee on-call interpretive performance on the subset of cases

from independent trainee call and has demonstrated capabilities in establishing benchmarks for major and minor discrepancy rates among trainees in a given institution [13]. This software is extremely useful for detecting errors of a trainee's interpretation of examinations during independent on-call experiences, but does not seek to evaluate the longitudinal reporting skills of trainees in all practice settings.

Another educational tool, the Report Comparator, is a web-based application that allows trainees and program directors to evaluate preliminary and finalized reports in a side-by-side fashion, as well as in a track changes mode [14]. This software allows for easy identification and visualization of changes among large numbers of reports dictated by specific trainees. The Report Comparator interface is designed to color code increasing numbers of changes made to displayed reports to draw trainee's attention to reports with more changes for further review, but it does not provide for quantitative analysis of reporting skills among trainees.

In this manuscript, we present the QRSE methodology to quantitatively evaluate all reports written by trainees during their residency experience. QRSE identifies specific trainees with significantly more and significantly less edits made to their preliminary reports during report finalization, thus benchmarking the quantity of edits made to trainee preliminary reports among residents over time. Furthermore, since all of the report metadata are stored in our database, trainee reporting skills information can be instantly evaluated by imaging modality, radiological subspecialty, time of day, and various other parameters.

This study establishes variance among trainees in the extent of changes made by attending radiologists during

the report finalization process. Of the trainees, 14.3 % had trainee-specific ALPs more than 2 standard deviations above the mean indicating that attending radiologists consistently made more changes to their reports than were made to other trainees. These more extensive changes may reflect a lower level of overall reporting skills. As a result of the QRSE, more focused efforts can be directed toward these trainees who may be struggling in their report generation.

Of the trainees, 16.7 % were identified who had trainee-specific ALPs less than 1 standard deviation below the mean, representing markedly fewer changes than average made to their preliminary reports during report finalization. Trainees with fewer edits made to their reports when working alongside a wide range of attending radiologists may have drafted high-quality radiology reports that closely mirror the report that would be issued by a more senior attending radiologist. These trainees may possess a high level of reporting skills and may also have a high level of radiological understanding that they leveraged in their report creation. The QRSE methodology identifies these trainees and can thus allow for them to receive positive reinforcement for their high reporting skills.

A possible limitation of this study is that a radiologist did not individually review the collective series of 228,543 radiology report pairs to evaluate the clinical significance of edits. However, a previous manual review of report edits made during report finalization showed that the quantity of edits positively correlated with the clinical significance of the edits made during report finalization, validating the quantity of edits as being a relevant metric for clinical significance of report changes [15].

Additional factors specific to attending radiologists may also play a role in determining the quantity of edits made to trainee's preliminary reports. One such factor could be the extent of discussion initiated by an attending radiologist about a radiology report prior to preliminary report dictation. More in-depth discussion of the findings of an examination may facilitate fewer changes being made during report finalization, but this is not known and remains a topic for further investigation. Furthermore, some attending radiologists may routinely make more or less edits than their peers. However, attending specific factors are applied across the board to trainees in a similar fashion and thus would tend not to affect individual trainee overall performance when using a metric such as the one we have suggested.

The relative number and type of cases a trainee interprets was also not formally standardized. It is possible that more edits are made on complex and lengthy CT and MR reports than on radiograph reports. However, because our institution schedules every trainee to experience the same weekly, biweekly, and/or monthly rotation changes among a fixed list of required rotations, we expect that given the extended

time period of the study, bias from factors related to the type of examination would be minimized.

The Accreditation Council for Graduate Medical Education (ACGME) has emphasized that medical residency training programs must develop systems, strategies, and tools to evaluate trainees on the ACGME Core Competencies [16]. These core competencies include: Patient Care, Medical Knowledge, Interpersonal and Communication, Professionalism, Practice-Based Learning Improvement, and Systems-Based Practice. Rigorous evaluation of the quality of trainee preliminary reports followed by targeted report improvement strategies for specific trainees furthers the mission of many of the six ACGME competencies [17].

This study proposes a QRSE methodology for the quantitative analysis of trainee-created preliminary reports. It further documents wide variation of the quantity of report edits made by attending radiologists to preliminary reports created by individual trainee radiologists in a large diagnostic radiology training program.

Further investigation may be useful to evaluate the QRSE methodology as a means to provide quantitative performance evaluation of individual trainee reporting skills, to monitor improvement during residency, and to evaluate the reporting skills of each class of trainees and within each radiological subspecialty.

Conclusion

The QRSE methodology longitudinally analyzes the quantity of report edits made to trainee reports by attending radiologists and assists in monitoring the reporting skills of individual diagnostic radiology trainees.

References

1. American College of Radiology: ACR standard for communication: diagnostic radiology. In: Standards. Reston: American College of Radiology, 1999–2000, pp. 1–3
2. Goodman NW, Edwards MB: Medical Writing: A Prescription for Clarity. Cambridge Univ. Press, New York, 1991
3. Hall FM: Language of the radiology report: primer for residents and wayward radiologists. *Am J Roentgenol* 175:1239–1242, 2000
4. Hunter TB: Radiographic reports: structure and review. (letter). *Am J Roentgenol* 142:647–648, 1984
5. American College of Radiology: The American College of Radiology DXIT™ & TXIT™ in-training examinations. Downloaded from <http://www.acr.org/Education/Exams-Certifications/DXIT-TXIT> on July 22, 2012
6. American Board of Radiology: The American Board of Radiology initial certification process. Downloaded from <http://www.theabr.org/ic-landing> on July 22, 2012
7. Google-diff-match-patch. Downloaded from <http://code.google.com/p/google-diff-match-patch/> on June 1, 2011

8. Carney E, Kempf J, DeCarvalho V, Yudd A, Noshier J: Preliminary interpretations of after-hours CT and sonography by radiology residents versus final interpretations by body imaging radiologists at a level 1 trauma center. *AJR Am J Roentgenol* 181:367–373, 2003
9. Wysoki MG, Nassar CJ, Koenigsberg RA, Novelline RA, Faro SH, Faerber EN: Head trauma: CT scan interpretation by radiology residents versus staff radiologists. *Radiology* 208:125–128, 1998
10. Tieng N, Grinberg D, Li SF: Discrepancies in interpretation of ED body computed tomographic scans by radiology residents. *Am J Emerg Med* 25:45–48, 2007
11. Meyer RE, Nickerson JP, Burbank HN, Alsofrom GF, Linnell GJ, Filippi CG: Discrepancy rates of on-call radiology residents interpretations of CT angiography studies of the neck and circle of Willis. *AJR Am J Roentgenol* 193:527–532, 2009
12. Itri JN, Kim W, Scanlon MH: Orion: a web-based application designed to monitor resident and fellow performance on-call. *J Digit Imaging* 24(5):897–907, 2011
13. Itri JN, Redfern RO, Scanlon MH: Using a web-based application to enhance resident training and improve performance on-call. *Acad Radiol* 17(7):917–920, 2009
14. Sharpe Jr, RE, Surrey D, Gorniak RJ, Nazarian L, Rao VM, Flanders AE: Radiology report comparator: a novel method to augment resident education. *J Digit Imaging* 25(3):330–336, 2012
15. Baghdanian A, Sharpe RE, Gorniak RJ, Nazarian LN, Rao VA, Flanders A: Quantitative metric to evaluate the clinical significance of changes made to trainee radiology reports. 2012 Annual Meeting of the Association of University Radiologists. March 21, 2012 (SS01-01). Downloaded from http://www.aur.org/Annual_Meeting/upload/Abstracts.pdf on July 22, 2012
16. ACGME: ACGME core competencies. Downloaded from http://www.acgme.org/acwebsite/RRC_280/280_corecomp.asp on June 1, 2012
17. Collins J, Rosado de Christenson M, Gray L, Hyde C, Koeller KK, Laine F, et al: General competencies in radiology training: definitions, skills, education and assessment. *Academic Radiology* 9(6): L721–L726, 2002