Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
ORIGINAL ARTICLE

Radiology Extenders: Impact on Throughput and Accuracy for Routine Chest Radiographs

Arijitt Borthakur, MBA, PhD, Eduardo M. Barbosa Jr, MD, Sharyn Katz, MD, Friedrich D. Knollmann, MD, PhD, Charles E. Kahn Jr, MD, MS, Mitchell D. Schnall, MD, PhD, Harold Litt, MD, PhD

Abstract

Interpretation of chest radiographs, although clinically highly impactful, entails a disproportionate amount of work relative to low reimbursement for these studies. Increasing volumes of chest radiographs in the thoracic imaging division of an academic radiology practice led to incorporation of two radiology extenders (REs) to draft reports of chest radiographs. We evaluated the difference in productivity and accuracy of the reports drafted by REs versus radiology trainees. Impact on throughput was analyzed by measuring flow rates (number of radiograph interpretations finalized per hour) for four subspecialist attending thoracic radiologists under three conditions: independent interpretation, reviewing RE-drafted cases, and reviewing resident-drafted cases. Improvements were calculated as change in flow rates for the latter two conditions compared with an independent interpretation. Accuracy of RE-drafted reports was compared with that of junior residents by evaluating their draft reports for the same chest radiographs (n = 49). A blinded judging panel of three attending radiologists scored these reports using the 3-point RADPEER scoring system; the original report dictated by an attending radiologist served as the reference standard. RADPEER scores were compared between the REs and residents. Flow rates improved significantly (P < .05) by 52% for RE-drafted cases versus a small improvement (17%) for resident-drafted cases compared with independently completed cases. RE-drafted cases generated 36% greater efficiency for attending radiologists compared with resident-drafted cases (P < .05). There was no significant difference in accuracy scores between RE-drafted and resident-drafted reports.

RE-drafted reports were finalized more rapidly by attending radiologists with insignificant differences in discrepancy rates compared with resident-drafted reports.

Key Words: Discrepancy rate, process improvement, radiograph, radiology extender

INTRODUCTION

Thoracic radiographs are performed in patients in the intensive care unit primarily to identify placement of support tubes, catheters, and other devices. Although highly impactful toward patient care, interpreting these radiographs entails a disproportionate amount of work (eg, retrieving patient history, completing standard dictation templates, and ensuring proper communication of important findings before finalization of reports). Given low reimbursement rates for these studies, economic necessities push radiologists to provide faster interpretations, contributing to burnout. Secondly, tasks that are characterized by low complexity and variability result in mundane work and a low sense of personal accomplishment, thus reducing job satisfaction for attending radiologists. Although both high throughput and repetitive tasks lead to underperformance in any occupation [1], including medicine [2,3], it has been particularly of concern in the field of radiology [4-6].

We previously described method to improve radiologists’ productivity by using a radiology extender (RE) in a musculoskeletal radiology section [7]. That experience motivated a trial of this concept in the department’s

Department of Radiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania.
Corresponding author and reprints: Arijitt Borthakur, MBA, PhD, Perelman School of Medicine at the University of Pennsylvania, Department of Radiology, 422 Curie Blvd, B1 Stellar-Chance Labs, Philadelphia, PA 19104-6100; e-mail: ari.borthakur@pennmedicine.upenn.edu.
Dr Mortani Barbosa Jr reports grants from Siemens AG, outside the submitted work. Dr Katz is a consultant for Trizell LTD (Imaging consultant for protocol development in a clinical trial of immunotherapy in malignant pleural mesothelioma). Dr Litt reports grants from Siemens Healthineers, outside the submitted work. The other authors state that they have no conflict of interest related to the material discussed in this article. Dr Borthakur, Dr Mortani Barbosa Jr, Dr Katz, Dr Knollmann, Dr Kahn, Dr Schnall, and Dr Litt are nonpartner, non–partnership track employees at the University of Pennsylvania.

J Am Coll Radiol 2020;[ ] Copyright © 2020 American College of Radiology

https://doi.org/10.1016/j.jacr.2020.09.044

Copyright © 2020 American College of Radiology

1546-1440/20/$36.00 https://doi.org/10.1016/j.jacr.2020.09.044

FLA 5.6.0 DTD ■ JACR5376_proof ■ 13 October 2020 ■ 1:28 pm ■ ce
cardiothoracic imaging division at our academic medical institute, which reports over 2,500 radiographs per week. The first experiment was to determine the impact of RE on throughput by measuring the time taken by radiology attending radiologists to finalize examinations. The second experiment compared the accuracy of RE-drafted reports to those from residents by comparing their dictation texts for the same chest radiographs. Double reading of imaging examinations is an accepted method of assessing interpretation quality [8,9]. In this approach, examinations are submitted to different reader(s) after the initial reader and are reinterpreted with the same clinical information and prior examinations that were available at the time of the initial interpretation by an attending radiologist. A third set of readers then compared both RE and resident-drafted reports and scored any discrepancies using the RADPEER system [10,11]. Major discrepancy rates were obtained by quantifying the highest score (3b) across REs and residents.

**METHODS**

**Recruiting and Training**
Both REs were recruited from an experienced pool of radiologic technologists through a formal interview process. Interview candidates responded to an advertisement circulated internally in our hospital system. Interviews were conducted over a period of a month by two radiology faculty, a clinical administration leader, and an operational business advisor. Nearly all candidates exhibited desirable characteristics such as radiographic technical knowledge, intellectual curiosity, and ability to decide under ambiguity. However, the two best candidates additionally had concrete evidence of performing beyond their current job requirements in completing tasks. Both REs were trained over a period of 2 months to interpret one- or two-view chest radiographs by a senior thoracic imaging radiologist with over 30 years of experience as a faculty member. Commensurate with previous RE training efforts [7], these included one-on-one review of cases predrafted by the trainee and directed reading in both an online course designed to train radiographers in the UK to interpret images (http://eintegrity.org, Ware, United Kingdom), as well as standard radiologic texts. Finally, the REs were trained to dictate using standardized templates in our hospital’s clinical reading rooms using standard reporting technology (ID37 PACS, Sectra Inc, Shelton, Connecticut; PowerScribe 360 reporting system, Nuance, Burlington, Massachusetts). The REs were deployed to the chest radiology section and drafted on average 130 to 160 cases each per day for a 6-month period. This project was undertaken as a quality improvement initiative and thus was exempt from institutional review board review.

**Flow Rate Measurements**
Four board-certified subspecialty-trained thoracic radiologists with at least 10 years of experience as faculty volunteered for this study. All radiographs were read in one of our hospital’s standard clinical reading rooms using the same reporting system utilized for training the REs. Examinations types were restricted to those routinely processed by the REs as part of their normal workflow. These consisted of chest one-view anterior-posterior or two-view lateral and posterior-anterior radiographs corresponding to Current Procedural Terminology codes 71010 or 71020, respectively. Examinations were not controlled for case variety or complexity, to replicate the realistic clinical scenario of the daily workflow. A single observer (A.B.) manually measured the time from examination opening in PACS to final signing in three categories: independently interpreted by the attending radiologist, RE-drafted reports, or resident-drafted reports. To reduce variability due to repeated stop-start events, the attending radiologists read out blocks of 5 to 30 or more cases under each category during each timing episode without interruption. These experiments were performed over several days to accumulate at least 50 cases for each category for each attending radiologist. Flow rate was defined as the number of radiographic examinations finalized per hour.

**Report Accuracy Comparison**
A sample of 49 attending radiologist-only finalized examination reports with Current Procedural Terminology codes 71010 and 71020 were extracted from the radiology database mPower (Nuance Communications) to serve as the reference standard. Only accession numbers, the unique record number in the local Radiological Information System for each diagnostic imaging examination, were supplied to the reporters who proceeded to dictate new reports using standardized templates without viewing the original attending radiologist’s report. These reports were copied into a Word (Microsoft Corp, Redmond, Washington) document and securely e-mailed to a single researcher (A.B.) for preprocessing, which involved automatically extracting report elements using custom-built code in Python 3.6 programming language written in a Jupyter notebook (jupyter.org), an open-source web application interface. All 49 extracted examination reports were stored on a scoring template created using Excel (Microsoft Corp) and saved using a coded file name to blind the reporter’s identity.

Reporter’s scoring template files were sent to a judging panel consisting of three attending radiologists with over 10 years of specialist thoracic imaging experience each. Judges were also provided the original attending radiologist reports to use as a gold standard. Comparisons were restricted to text of the reports and no images were supplied. Diagnostic discrepancies were classified according to ACR’s 2016...
262 RADPEER scoring system [11]. This scoring system 
determines the significance of a discrepancy based on a 3-
point scale: 1 = concur with interpretation, 2 = discrep-
ancy in interpretation or not ordinarily expected to be made 
(understandable miss), and 3 = discrepancy in interpreta-
tion or should be made most of the time. Modifiers for 
scores 2 and 3 include a = unlikely to be clinically signifi-
cant and b = likely to be clinically significant, resulting in 
five possible scores. RADPEER scores of 3b were deemed a 
major discrepancy and their count was recorded in the Excel 
spreadsheet for both REs and residents.

Statistical Analysis

Statistical analyses were performed in JMP Pro 15 (SAS, 
Cary, North Carolina) and results graphed in Excel. A P 
value of <.05 was assumed to be statistically significant for 
all hypothesis testing. For the flow rate measurements, 
average read time, defined as the time taken between 
opening the first case in a queue until finalization of the last 
case, was recorded within a fraction of a minute. Analysis 
of variance was performed on flow rate measurements with 
three-level categorical dependent variables (ie, indepen-
dently interpreted, resident, and RE), resulting in a t test 
with pooled standard error. An ordered logit model [12] was 
used to regress ordinal RADPEER scores against the 
reporter category (ie, RE or resident) and accession 
number of the 49 cases to account for case mix. The model fits 
the cumulative response probabilities to the 
logistic function of a linear model using maximum 
likelihood. The value of the likelihood ratio \( \chi^2 \) statistic 
and associated P value were recorded as the figure of merit 
for the corresponding effect. A low value of the statistic 
means that the null hypothesis should not be rejected.

RESULTS

The REs at our practice were formally recruited from a pool 
of x-ray radiologic technologists (rad techs or RTs), whose 
qualifications vary by state but must have at least an asso-
ciate’s degree or a technical diploma. Both of our REs have 
bachelor’s degrees in addition to being licensed radiologic 
technologists for at least 6 years. One studied art history 
in college and served as a licensed medic, and another has an 
undergraduate degree in medical imaging, completed her 
RT training at our institute, and served as the lead inpatient 
tech for the preceding 2 years.

Flow rate for the three categories of examinations 
finalized by the four attending radiologists that 
participated in the timing study are shown in Figure 1. RE-drafted cases 
were finalized more rapidly compared with resident-drafted 
cases as well as cases completed by the attending radiologists 
by 52% for RE-drafted cases compared with a smaller 
 improvement of 17% for resident-drafted cases, which was 
within the error variance of flow rate of independently 
completed cases. RE-drafted cases were also finalized 36% 
(P < .05) faster by the attending radiologists compared with 
resident-drafted cases, suggesting perhaps that more edits 
were necessary on resident compared with RE reports. 

The distribution of RADPEER scores across the re-
porters (Fig. 2) revealed no significant difference in scores 
between RE-drafted and resident-drafted reports. The ma-
jor discrepancy rate amounted to less than 2% of all cases 
and generated only by the REs. Modeling scores with re-
porter type and examination case mix (ie, accession number) 
showed a 100-fold greater contribution by the latter variable 
to the likelihood-ratio \( \chi^2 \) (Table 1), regardless of whether 
the reporter was a resident or an RE. Combining these 
results suggested that the variability in scores was 
significantly (P < .05) related to case mix and indifferent 
to reporter type.

DISCUSSION

The results of this project revealed baseline values for flow 
rate improvements, a key performance indicator for gauging
efficiency gains from incorporating an RE in practice. Previous work at our institute that graded trainee reports demonstrated that the rate of discrepancies as measured by major changes during attending radiologist review varied from 1.5% to 4% among individual residents, modalities, and level of resident training [13,14]. Hence the discrepancy rate of the REs is within that expected of trainees. Major discrepancy rates, although low and consistent with values measured previously at our institution, still merit a deeper evaluation to detect patterns of mischaracterizations or bias by the reporters. Measured values will be leveraged to optimize the deployment of REs into the reading rooms to synchronize their work schedules with fluctuating radiography volumes during the course of a day or a week. Furthermore, the variability in the efficiency gains across radiologists could be utilized to redesign the processes of allocating cases into each of their queues.

The RE job role is controversial among some who consider REs as a cost-effective replacement for radiology trainees. We respectfully disagree with their presumption. The incorporation of REs had been a strategic decision to address recent mergers and acquisitions that have resulted in our radiology service providing care in multiple affiliated hospitals and diagnostic imaging centers located across our region and growth within those facilities, performing over a million cases annually. As with other academic hospitals, we have relied on trainees to share the burden of interpreting studies at lower cost; however, their supply has not kept pace with the increasing volume of work. Regardless, as an academic medical center we hold training the next generation of radiology leaders at a higher priority than any revenue improving or cost control initiative. We are confident that the RE job function makes more business sense if REs are deployed in a strategic and meaningful manner.

Physician wellness is acknowledged as an essential goal in enhancing the safety and quality of health care [15]. Although we are in the process of surveying radiologists' satisfaction with the new workflow, concerns in response to our earlier work on the role of REs included the suggestion that reallocation of radiologists' tasks would devalue their work. We contend that the diagnostic responsibility remains with the radiologist in the RE-enhanced workflow while some of the technical tasks such as handling the computer software for speech recognition and text editing are taken over by support staff. Our model does not assign the intellectual contribution of the radiologist, but rather improves their professional stance by relieving them of some of the information technology tasks in a manner that fits the modern, fast-paced health care arena better than other methods such as tape-recording dictation models. We continue to manage the process improvement change proactively by tracking success with performance indicators such as average daily volumes by each RE. Over a typical month, each RE predrafts upwards of 160 cases per day, which at the activity time difference measured would free up 51 min of a radiologist's time per day for higher value work.

Our experience in transforming our radiology practice required effective change management approaches to clearly articulate the business need and projected impact clearly. We relied on data from a variety of internal systems such as portable x-ray, radiologist, and resident schedules to optimize the work hours for the REs and on external sources such as online courses for training. Although the current REs at our practice are licensed x-ray technicians with bachelor's degrees, as the role evolves and is adopted more

Table 1. Results of ordinal logistic modeling RADPEER score with explanatory variables reporter type and examination accession number

| Variable                          | DF | L-R $\chi^2$ | P Value |
|-----------------------------------|----|--------------|---------|
| Reporter type (radiology extender/resident) | 1  | 1.15         | .35     |
| Case mix                         | 48 | 92.9         | <.05    |

These results suggested that the variability in RADPEER scores was significantly ($P < .05$) related to case mix (ie, examination accession number) and indifferent to reporter type. DF = degrees of freedom, L-R $\chi^2$ = likelihood ratio $\chi^2$ statistic.
widely, we foresee additional licensing and certification requirements, most likely on a state-by-state basis.

A limitation of our approach was that an ideal quality evaluation should also assess the accuracy of the initial report, independently verifying those findings with operative findings, pathology reports, or a consensus interpretation by experts. This would lead to a better understanding of the discrepancies observed in the current study and will be pursued in a subsequent study. Another limitation for the timing experiments was that the differences in types of cases was not controlled as attending radiologists were timed finalizing cases each morning as part of their routine clinical workflow. There may also be a "Hawthorne effect," sometimes described as the increase in human productivity when people are observed and measured in real time [16], and may be different for each radiologist. Additionally, although these interpretations were performed as part of the radiologists’ clinical work, they were completed and timed without interruption, which may not reflect a typical routine.

As their medical colleagues have become more comfortable with their services, the REs have taken on additional responsibilities in the reading room. A frequent complaint of radiologists is the many daily interruptions that impede one’s ability to sustain a high rate of examination interpretations. Fishman et al [5] describe these interruptions to include telephone calls with referring physicians to communicate important results and working with technologists regarding protocoling or image acquisition. A recent publication from our institution on deploying a “call triage assistant” demonstrated improved workflow efficiency and reduced resident stress on call [17]. At our practice, the REs are routinely becoming the first to answer telephone calls and communicate previously reported findings to referring physicians, routing more difficult questions to senior house staff or attending radiologists as needed.

Finally, advances in artificial intelligence and machine learning techniques, particularly deep learning algorithms, have demonstrated significant success in image recognition [18]. In thoracic imaging, these algorithms have been able to perform a variety of tasks such as automatically identifying and categorizing benign or malignant lung nodules in CT images, predicting long-term mortality from chest radiographs [19], and distinguishing COVID-19 from other types of pneumonia from CT images [20]. As artificial intelligence becomes more integral to patient care in radiology, radiologists who have leveraged its potential will likely become indispensable to the field [21]. We foresee the RE role evolving in conjunction to serve as a data curator, leveraging their technical expertise to assess image quality and guide reacquisition and performing hyperparameter tuning during model training to accelerate the adoption of machine learning algorithms in radiology workflow.

**TAKE-HOME POINTS**

- Deployment of REs in the thoracic imaging section of our academic radiology practice improved efficiency of chest radiography.
- RE-drafted reports were finalized more rapidly than resident-drafted reports by attending radiologists with insignificant differences in interpretation discrepancy rates.

**ACKNOWLEDGEMENTS**

The authors thank Orli Gross B.S., R.T (R), for help retrieving studies and participating in their evaluation. They also thank Michelle Weigle, BS, RT (R), Hena Cheema, MD, Taylor Pomeranz, MD, and Scott Simpson, DO, for participating in evaluation of exams.

**REFERENCES**

1. Hackman JR, Greg O, Robert J, Kenneth P. A new strategy for job enrichment. Calif Manage Rev 1975;17:57-71.
2. Dyrybe LN, Awad KM, Fiscus LC, Sinsky CA, Shanafelt TD. Estimating the attributable cost of physician burnout in the United States. Ann Intern Med. United States 2019;600-1.
3. Shanafelt TD, West CP, Sinsky C, et al. Changes in burnout and work-life integration in physicians and the general US working population between 2011 and 2017. Mayo Clinic Proc 2019;94:1681-94.
4. Kruskal JB, Norbash A. A call to action—our radiology chairs are burning out. Acad Radiol 2019;26:1385-9.
5. Fishman MDC, Mhta TS, Siewert B, Bender CE, Kruskal JB. The road to wellness: engagement strategies to help radiologists achieve joy at work. Radiographics 2018;38:1651-64.
6. Chetlen AL, Chan TL, Ballard DH, et al. Addressing burnout in radiologists. Acad Radiol 2019;26:526-33.
7. Borthakur A, Kneeland JB, Schnall MD. Improving performance by using a radiology extender. J Am Coll Radiol 2018;15:1300-3.
8. Sahni VA, Khorasani R. Radiology double reads. BMJ Qual Saf 2016;25:569-71.
9. Geijer H, Geijer M. Added value of double reading in diagnostic radiology, a systematic review. Insights Imaging 2018;9:287-301.
10. Geijer H, Geijer M. Added value of double reading in diagnostic radiology: a systematic review. Insights Imaging 2018;9:287-301.
11. Goldberger-Stein S, Frigini LA, Long S, et al. ACR RADPEER Committee white paper with 2016 updates: revised scoring system, new classifications, self-review, and subspecialized reports. J Am Coll Radiol 2017;14:1080-6.
12. McCullagh P. Regression models for ordinal data. J Roy Stat Soc Series B Methodol 1980;42:109-42.
13. Rattoiiatinen 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;8:644-8.
14. Itri JN, Kim W, Scanlon MH. Orion: a web-based application designed to monitor resident and fellow performance on-call. J Digit Imaging 2011;24:897-907.
15. Dewa CS, Loong D, Bonato S, Trojanowski L. The relationship between physician burnout and quality of healthcare in terms of safety and acceptability: a systematic review. BMJ Open 2017;7: e015141.

16. Kidwai AS, Abujudeh HH. Radiologist productivity increases with real-time monitoring: the Hawthorne Effect. J Am Coll Radiol 2015;12:1151-4.

17. Levy JL, Freeman CW, Cho JK, Iyalomhe O, Scanlon MH. Evaluating the impact of a call triage assistant on resident efficiency, errors, and stress. J Am Coll Radiol 2020;17:414-20.

18. Choy G, Khalilzadeh O, Michalski M, et al. Current applications and future impact of machine learning in radiology. Radiology 2018;288:318-28.

19. Lu MT, Ivanov A, Mayrhofer T, Hosny A, Aerts HJWL, Hoffmann U. Deep learning to assess long-term mortality from chest radiographs. JAMA Netw Open 2019;2:e197416.

20. Li L, Qin L, Xu Z, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. Radiology 2020;296:E65-71.

21. Chan S, Siegel EL. Will machine learning end the viability of radiology as a thriving medical specialty? Br J Radiol 2019;92:20180416.