

Radiologists' Variation of Time to Read Across Different Procedure Types

Daniel Forsberg^{1,2} \bullet Beverly Rosipko² \cdot Jeffrey L. Sunshine²

Published online: 6 October 2016 \oslash Society for Imaging Informatics in Medicine 2016

Abstract The workload of US radiologists has increased over the past two decades as measured through total annual relative value units (RVUs). This increase in RVUs generated suggests that radiologists' productivity has increased. However, true productivity (output unit per input unit; RVU per time) is at large unknown since actual time required to interpret and report a case is rarely recorded. In this study, we analyzed how the time to read a case varies between radiologists over a set of different procedure types by retrospectively extracting reading times from PACS usage logs. Specifically, we tested two hypotheses that; i) relative variation in time to read per procedure type increases as the median time to read a procedure type increases, and ii) relative rankings in terms of median reading speed for individual radiologists are consistent across different procedure types. The results that, i) a correlation of -0.25 between the coefficient of variation and median time to read and ii) that only 12 out of 46 radiologists had consistent rankings in terms of time to read across different procedure types, show both hypotheses to be without support. The results show that workload distribution will not follow any general rule for a radiologist across all procedures or a general rule for a specific procedure across many readers. Rather the findings

 \boxtimes Daniel Forsberg daniel.forsberg@sectra.com

> Beverly Rosipko beverly.rosipko@uhhospitals.org

> Jeffrey L. Sunshine jeffrey.sunshine@uhhospitals.org

- ¹ Sectra, Teknikringen 20, SE-583 30 Linköping, Sweden
- ² Department of Radiology, Case Western Reserve University and University Hospitals Cleveland Medical Center, 11100 Euclid Avenue, Cleveland, OH 44106, USA

suggest that improved overall practice efficiency can be achieved only by taking into account radiologists' individual productivity per procedure type when distributing unread cases.

Keywords PACS · Radiology workflow · Efficiency · Productivity

Introduction

A radiologist's workload is commonly assessed by determining the total annual relative value units (RVUs) [\[1](#page-7-0), [2](#page-8-0)], where an RVU for a specific procedure is considered to provide a quantitative metric describing how much work is associated with performing a specific procedure in comparison with other procedure types. As a workload metric, RVUs are considered far better than simply counting the number of performed procedures as different procedures can be more or less consuming and/or complex to interpret and report on for a radiologist. Over the past few decades, it has been frequently noted that the total annual RVUs or the adjusted RVUs per full-time equivalents have increased [\[3](#page-8-0)–[6](#page-8-0)]. These data are often interpreted to suggest that not only workload but productivity has increased. However, there is actually very little knowledge about true productivity (defined as output unit per input unit) for radiologists interpreting and reporting on radiological examinations. This is in part because the work carried out by a radiologist includes more than merely the RVU generating activities of interpreting and reporting (for example, patient contact, communication with colleagues, teaching, research, and administrative duties). In addition, there are large variations between radiology groups and subspecialties [\[7](#page-8-0)–[10\]](#page-8-0), and most notably time spent on interpreting and reporting a case is at large unknown. A few exceptions can be found, for example in [\[11\]](#page-8-0), where measured reporting times was used to determine actual RVUs for a set of procedure types, or in [[12\]](#page-8-0), where the work of three radiologists was closely monitored and a low overall correlation between time to read and RVU was found.

It might be argued that actual productivity is of limited relevance as long as a radiologist delivers as anticipated, i.e., whether it takes 1 h to read 30 chest radiographs or 2 h is less relevant as long as the task is successfully completed within a certain time frame, as for example specified through a radiology group's service level agreement. However, in times when volumes are increasing and resources are limited, productivity and the optimization thereof play an increasingly pivotal role. Especially if, for example, a set of unread cases can be distributed among a group of radiologists to increase overall efficiency [[13\]](#page-8-0) or where the practice of a highly productive radiologist can be mimicked by a less productive radiologist to increase efficiency [\[14\]](#page-8-0). Hence, knowledge of productivity, factors affecting it and methods to optimize overall efficiency are of rising importance. That time to read a case indeed depends on both radiologist and procedure type was shown in [\[15\]](#page-8-0) but to what extent these and how additional factors such as age, years of experience and subspecialty affect productivity is at large unknown.

Our primary interest in these questions stem from an ongoing work at our institution focused on how to efficiently distribute unread cases among radiologists. The following alternatives can be considered to reflect two opposite alternatives for work distribution. The first alternative is to have a single worklist with all unread cases from which essentially all radiologists can select any case at their own discretion. The second (opposite) alternative is to have no worklists at all and where instead each individual radiologist is automatically fed with each next case to read (based upon each radiologist's subspecialty, earlier productivity and quality, and the group's overall workload) without an opportunity to themselves influence what to read next. A first step in understanding the potential effects of different options for workload distribution is to gain a better understanding as how time to read varies over a set of procedure types and between radiologists.

The main purpose of this study is to determine how the coefficient of variation (CV) of time to read, and how the relative ranking among radiologists (based on time needed to read) vary between procedure types. Based upon our own experience and assumptions, we hypothesize (i) that the relative variation in time to read a procedure type will be positively correlated with the median time to read that procedure type (more variation the longer time required to read a certain procedure type) and (ii) that the relative ranking among radiologists in productivity (time to read) is consistent over different procedure types for most radiologists (a radiologist is, for example, always fast or slow regardless of procedure type).

Material and Methods

This study was considered as exempt from review by our institutional review board as part of a larger quality project in our department.

Overview of the Radiology Reading Environment

The study was conducted at an institution comprising a health system anchored by a tertiary and quaternary academic medical center comprised of four hospitals including their satellite outpatient centers, with additional eleven community hospitals in the surrounding geography. The radiologist serving the health system is split into two groups. One group primarily serving the academic medical center and the other group serving essentially all the community hospitals. The radiologists at the academic medical center are all involved in teaching activities and only read according to their respective subspecialty (abdominal imaging, breast imaging, neuroradiology, musculoskeletal, nuclear medicine, thoracic radiology and vascular/interventional radiology). The community radiologists, on the other hand, have no teaching responsibilities and read subspecialized to a lesser degree.

The radiologists work in a PACS-driven workflow, where a PACS workstation has a number of worklists from which the radiologists select what case to read next. The worklists are typically populated depending on examination status, location, modality and body part. Each radiologist has a number of specified worklists to cover each day according to a given schedule. The academic radiologists only cover worklists according to their respective subspecialty whereas the community radiologists can cover multiple subspecialties. All radiologists use voice recognition (a separate system but integrated with the PACS) for dictation of reports.

Data and Pre-processing

The PACS (Sectra PACS 17.3, Sectra North America Inc., Shelton, CT, USA) installed at our health system provides the possibility to log all commands issued by each PACS user. The logging can be controlled to determine which users, applications and/or workstations to log commands for. The log files themselves are stored as standard text files, although with a specific formatting, and contain information such as which command, who issued the command, when and on which workstation. In the current study, command logs from May 4th to August 30th 2015 were extracted for processing, i.e., 17 weeks of data was extracted covering data from residents, fellows and attending radiologists. This data served as input to extract information about radiologist, procedure type and time to read for procedures performed between May 4th and August 23rd 2015 (16 weeks).

Before analysis, the logs were pre-processed. This consisted of concatenation of the daily logs to a single file, splitting of user sessions into separate traces for each opened examination (where each trace would record user, information about the opened examination and commands issued) and removal of commands irrelevant to the PACS reading workflow. Removed commands included, for example, commands for navigation (e.g., among worklists), search commands and login/logout commands. Table 1 provides an example command trace. Note that the command traces contain log entries for when a dictation is initiated from the PACS but actual commands (start, stop, etc.) for the dictation process are not available, since the dictation application is separate from the PACS. Next, each command trace was linked via the available examination information to actual examinations recorded in the PACS database, which allowed us to update each trace with its exact procedure type and to filter out command traces pertaining to examinations performed outside of the selected time frame (May 4th to August 23rd 2015, 16 weeks). In addition, traces that could not be linked to a single valid examination in the PACS database or that were related to interventional, vascular or nuclear medicine procedures were discarded. These procedure types were excluded since most of the radiologist's work associated with these takes place outside of the PACS. The initially extracted data set contained 459,263 command traces and 235,976 unique examination identifiers recorded from 155 radiologists and 51 residents. After the initial filtering, command traces corresponding to 162,857 examinations remained.

In the next step, each set of command traces belonging to the same examination were analyzed and command traces from examinations where only a single user during a single occasion had dictated and reviewed the examination were retained (discarding command traces corresponding to 67,071 examinations). This filtering step ensured, for example, that cases reviewed by both a trainee

Table 1 Example command trace from a radiologist reading a single view chest radiograph. In this example, apart from the start and end commands, only five different commands are issued during the review

User	007
Role	Radiologist
Examination information	$6/15/2015$, 11:54 AM, chest AP view
Date time	Command
2015-06-15T13:25:18	Set current exam
2015-06-15T13:26:25	Dictate report
2015-06-15T13:26:47	Pan
$2015 - 06 - 15T13.26.48$	Zoom
$2015 - 06 - 15T13.26.54$	Next hanging
2015-06-15T13:26:58	Previous hanging
2015-06-15T13:27:33	Artificial end

and an attending would be excluded. In addition, traces recording a time difference of X minutes or more between subsequent commands were excluded to avoid traces recording interrupted reading sessions (5 min for CT and MRI procedures and two for all other procedures). The longer time difference for volumetric studies was chosen to accommodate for potential prolonged time periods of just scrolling in an image series, an activity not recorded by the command logs. Further, only traces recorded by attending radiologists and with procedure types for which a radiologist had read at least 20 cases were retained. As a last filtering step, only data pertaining to procedure types read by at least three different radiologists and radiologists who have read at least three different procedure types was included for the subsequent data analysis. After the final filtering, only 57,881 command traces/ examinations from 46 radiologists covering 54 different procedure types remained.

For each trace radiologist, procedure type and time to read were extracted. The time to read was computed as the time difference between the time stamps of the first and last commands of each trace.

Statistical Analysis

For each procedure type, the median, mean and standard deviation (STD) of time to read were estimated, and from the two latter the CV was computed. The Pearson correlation was used to determine the linear relationship between the median time and CV of each procedure type.

In addition, the median times to read for each radiologist and procedure type were computed. These median times were used to compute a relative ranking of each radiologist per procedure type. This was done by first setting a rank r_{kl} for each radiologist k with rank 1 for the slowest and rank n_l for the quickest of procedure type *l*. The rankings were then normalized to the value range $(0,1)$.

$$
r_{kl}=1-\frac{r_{kl}}{n_l+1}
$$

For the normalized relative rankings, a higher value signifies a quicker reader.

The consistency of the relative rankings for each radiologist was tested with a one-sided non-parametric sign test with significance level of 0.05. The test was applied to determine if the variable $x_{kl} = abs(r_{kl} - median(\lbrace r_{k1}, r_{k2}, ..., r_{kl}\rbrace))$ for radiologist k had a distribution where the median was larger than 0.15. The threshold of 0.15 was empirically chosen and was considered to provide a reasonable comprise to allow some variation in the relative rankings but not too much.

Results

Figure 1 provides an overview of the results showing the median time for each radiologist and procedure type, where each marker (per procedure type signifies) a different radiologist. The procedure types are sorted according to their respective median times to read. An increased absolute dispersion of median times to read per radiologist and procedure type can be observed from left to right, which indicates support for our hypothesis that the variation in time to read increases with increased time to read per procedure type. For most radiologists, it appears as if the ranking between the radiologists varies quite a lot across the different procedure types.

The computed median, mean, STD and CV for each procedure type can be observed in Table [2.](#page-4-0) The results show our first hypothesis to be without support as the CV for the time to read actually decreases as the median time increases per procedure type. This is further supported by an estimated negative correlation of $r = -0.25$ (p value = 0.067) between the median and CV. If we analyze the data grouped per modality, the estimated correlations between the median time and CV per procedure type are $r = -0.42$ (p value = 0.021) for CR, $r = -0.86$ (p value = 0.006) for CT, $r = -0.02$ (p value = 0.969) for MRI and $r = -0.74$ (0.022) for US. Again, the results per modality show no support for our initial hypothesis. Figure [2](#page-5-0) shows a plot of the median time to read and the computed CV along with the fitted linear regression lines for all procedure types and per modality. Although the negative correlation between the median time to read and CV differed some between the modalities, the slopes of the fitted linear regression lines are very similar between CR, CT and US. The exception here is for the MR procedures, but which are very few and with little variation for both median time to read and CV. Note that for MG no correlation has been estimated and no linear regression has been performed, since only two MG procedure types were included. However, a line fitted to the two available procedure types would yield a slope similar to the other regression lines.

Other noteworthy results from Table [2](#page-4-0) include procedures types that use standard X-ray imaging of the thorax and/or the abdomen have higher CV compared to procedure types using standard X-ray of the skeleton; neurological MRI procedure types have relatively high CV values; thorax and abdominal CT procedure types have relatively low CV values. For the CT examinations, it can be noted that contrast enhanced procedures types have a slightly larger CV than the corresponding non-contrast enhanced procedure types.

Figures [3](#page-6-0) and [4](#page-6-0) show the results from analyzing the relative rankings for each radiologist's time to read per procedure type. Radiologists with rankings considered as consistent by the statistical test are visualized in Fig. [3](#page-6-0), whereas inconsistent

Fig. 1 Median time to read per procedure type and radiologist, sorted according to overall median time to read per procedure type. The thicker black line corresponds to the overall median time to read for all radiologists per procedure type

Table 2 Computed median, mean, standard deviation, and CV for time to read over all procedure types

Table 2 (continued)

radiologists are visualized in Fig. [4](#page-6-0). The results show that only 12 out of 46 radiologists have rankings considered to be consistent across different procedure types. Hence, most radiologists can be considered to vary significantly in their relative rankings as derived from their respective times to read over different procedure types, which leaves no support for the second hypothesis. Interestingly, radiologists that were found to have consistent relative rankings can still have certain procedure types for which they have an essentially opposite ranking, e.g., a radiologist that in general is a quick reader might be among the slowest for some procedure types. Of the included 46 radiologists, 15 were active at the academic medical center and, therefore, primarily read subspecialized. The results in Figs. [3](#page-6-0) and [4](#page-6-0), also show that 4 out of 15 subspecialized radiologists were consistent in their relative rankings, whereas 8 out of 34 general radiologists were consistent in their relative rankings. Hence, no apparent difference can be noted in terms of consistency between general and subspecialized radiologists.

Discussion

In this study, we analyzed how the time to read a case varies between radiologists over a set of different procedure types. Specifically, we set out to test the hypotheses (i) that the relative variation in time to read per procedure type increases as the median time to read a procedure type increases and (ii) that the relative rankings in terms of reading speed for individual radiologists are consistent across different procedure types. Results show, however, both hypotheses to be without statistical support.

In terms of the relative variation in time to read it was found that the median time and CV per procedure type had a

Fig. 2 Median time to read and CV per procedure type with fitted linear regressions for all procedure types and per modality. The different symbols correspond to the following modalities: \Diamond = CR, + = CT, x = MG, \circ = MR, \Box = US

Fig. 3 Relative rankings for time to read per procedure type and radiologist for radiologists considered to be consistent in their relative rankings. Dotted lines correspond to general radiologists and dashed to subspecialized

Fig. 4 Relative rankings for time to read per procedure type and radiologist for radiologists considered to be inconsistent in their relative rankings. Dotted lines correspond to general radiologists and dashed to subspecialized

negative correlation of −0.25. A potential explanation for this result could be that procedure types that require longer times to read are more complex and, therefore, demands a much more careful interpretation process to be carried out by all radiologists, whereas the interpretation process of a more "simple" procedure type can be done a great deal more efficiently by experienced radiologists.

Note that the difference in results as observed in Fig. [1](#page-3-0) and found in Table [2](#page-4-0) is to some extent due to only median times being displayed in Fig. [1](#page-3-0). The use of only median times limits the influence of any outliers in the data, which on the other hand has a potentially substantial impact on the mean, the standard deviation and the subsequently computed CV. Consequently, the reliability of the correlation between the median time and CV is somewhat impeded because of the large difference between the median and the mean, indicating that the data contains a large number of outliers. Regardless of this, note the overall large values of the CV for most procedure types are considerably large. Hence, it is relevant to further inquire whether reading workflow standardization would be of relevance and could potentially decrease the average time to read.

Out of 46 included radiologists, only 12 were found to have consistent rankings of time to read across the set of included procedure types. An important practical implication of this result is that since the relative rankings vary across different procedure types it would make sense to distribute workload according to radiologists' previously recorded performance. For example, a radiologist who is quick reading 'DIGITAL MAMMOGRAPHY SCREENING' might preferably read those but avoid procedure types where his or her performance is significantly slower than others (assuming all other factors including quality remain consistent). This is somewhat exemplified in the results provided in Fig. [3](#page-6-0), as some of the radiologist marked as consistent in their relative rankings still have some large variations in their relative rankings. The results, in Figs. [3](#page-6-0) and [4,](#page-6-0) also showed that there was no apparent difference for the consistency of relative rankings between general and subspecialized radiologists, and with both general and subspecialized radiologists found to be quick respectively slow readers.

The generalizability of the presented study is foremost limited by the small data set size and that data from only a single radiology group has been considered, although including both general and subspecialized radiologists. The small data set size was in part due to a rather conservative approach for filtering the data employed, for example, only including data where a single radiologist at a single occasion had interpreted and reported on a case. This approach might have a bias of excluding more complex cases where a radiologist performs an initial read and then later returns to the case after a consultation with a colleague. In addition, a lot of the work performed by the radiologists at the academic medical center was excluded because of their involvement in teaching,

causing only a rather low number of subspecialized radiologists to be included in the study. A specific limitation for US data is that the measured time might include cases where the interpreting radiologist performed the scanning as well and, thus, has already reviewed the images to some extent before the actual review in the PACS. Another important limitation is that the time to read is estimated from command logs recording PACS usage where there does not exist a specific command to denote that a reading is completed. Instead, time to read was computed as the time difference between the time stamps of the first and last commands of each trace, recording all the issued commands for an opened examination. Finally, we have no measure of quality of interpretation such as accuracy of findings, weight of impressions, or consistency of follow up to accepted standards.

Conclusion

With the analysis of command logs recording radiologists PACS usage, we were able to determine time to read per case, which in turn allowed us to perform an in-depth analysis of how time to read varies between radiologists and over different procedure types. We found that the degree of relative variation for time to read, as given by the coefficient of variation, to some extent decreases as the median time to read per procedure type increases. Further, results showed that relatively few radiologists are consistent in their rankings in terms of time to read across different procedure types. Both of these findings have implications for workload distribution, especially the latter, suggesting that there may be productivity opportunity for improvements by routing cases based upon procedure types to individual radiologist worklists instead of to generic examination work lists (regardless if the radiologist is a general or subspecialized radiologist). However, a better understanding of how a worklist-less work situation will affect a radiologist is needed before implementing such an approach. In addition, these results suggest one may further review the process for reading examinations with a large degree of variation to see if factors that cause this variation can be identified.

Compliance with Ethical Standards

Funding D. Forsberg is supported by a grant (2014-01422) from the Swedish Innovation Agency (VINNOVA).

References

1. Hsiao WC, Braun P, Becker ER, Thomas SR: The resource-based relative value scale: toward the development of an alternative physician payment system. JAMA 258:799–802, 1987

- 2. Conoley PM, Vernon SW: Productivity of radiologists: estimates based on analysis of relative value units. AJR Am J Roentgenol 157:1337–1340, 1991
- 3. Conoley PM: Productivity of radiologists in 1997: estimates based on analysis of resource-based relative value units. AJR Am J Roentgenol 175:591–595, 2000
- 4. Arenson RL, Lu Y, Elliott SC, Jovais C, Avrin DE: Measuring the academic radiologist's clinical productivity: survey results for subspecialty sections. Acad Radiol 8:524–532, 2001
- 5. Lu Y, Zhao S, Chu PW, Arenson RL: An update survey of academic radiologists' clinical productivity. J Am Coll Radiol 31:817–26, 2008
- 6. Bhargavan M, Kaye AH, Forman HP, Sunshine JH: Workload of Radiologists in United States in 2006–2007 and Trends Since 1991–1992 1. Radiology 252:458–67, 2009
- 7. Sunshine JH, Burkhardt JH: Radiology Groups' Workload in Relative Value Units and Factors Affecting It. Radiology 214: 815–822, 2000
- 8. Eschelman DJ, Sullivan KL, Parker L, Levin DC: The relationship of clinical and academic productivity in a university hospital radiology department. AJR Am J Roentgenol 174:27–31, 2000
- 9. Duszak R, Muroff LR: Measuring and managing radiologist productivity, part 1: clinical metrics and benchmarks. J Am Coll Radiol 7:452–458, 2010
- 10. Duszak R, Muroff LR: Measuring and managing radiologist productivity, part 2: beyond the clinical numbers. J Am Coll Radiol 7: 482–489, 2010
- 11. Cowan IA, MacDonald SL, Floyd RA: Measuring and managing radiologist workload: Measuring radiologist reporting times using data from a Radiology Information System. J Med Imaging Radiat Oncol 57:558–566, 2013
- 12. Krupinski EA, MacKinnon L, Hasselbach K, Taljanovic M: Evaluating RVUs as a measure of workload for use in assessing fatigue. SPIE Proc 9416:94161A, 2015
- 13. Halsted MJ, Froehle CM: Design, implementation, and assessment of a radiology workflow management system. AJR Am J Roentgenol 191:321–327, 2008
- 14. Reiner B: Automating radiologist workflow, part 2: hands-free navigation. J Am Coll Radiol 30:1137–41, 2008
- 15. Forsberg D, Rosipko B, Sunshine JL: Factors Affecting Radiologist's PACS Usage. J Digit Imaging, 2016 (EPub)