Work overload and diagnostic errors in radiology

Kasalak, Ömer; Alnahwi, Haider; Toxopeus, Romy; Pennings, Jan P; Yakar, Derya; Kwee, Thomas C

Published in:
European Journal of Radiology

DOI:
10.1016/j.ejrad.2023.111032

Publication date:
2023

European Journal of Radiology 167 (2023) 111032

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Work overload and diagnostic errors in radiology

Ömer Kasalak *, Haider Alnahwi, Romy Toxopeus, Jan P. Pennings, Derya Yakar, Thomas C. Kwee

Medical Imaging Center, Department of Radiology, University Medical Center Groningen, University of Groningen, the Netherlands

ARTICLE INFO

Keywords:
Computed Tomography
Diagnostic Errors
Workload

ABSTRACT

Purpose: To determine the association between workload and diagnostic errors on clinical CT scans.

Method: This retrospective study was performed at a tertiary care center and covered the period from January 2020 to March 2023. All clinical CT scans that contained an addendum describing a perceptual error (i.e. failure to detect an important abnormality) in the original report that was issued on office days between 7.30 a.m. and 18.00 p.m., were included. The workload of the involved radiologist on the day of the diagnostic error was calculated in terms of relative value units, and normalized for the known average daily production of each individual radiologist ($workload_{normalized}$). A $workload_{normalized}$ of less than 100% indicates relative work underload, while a $workload_{normalized}$ of > 100% indicates relative work overload in terms of reported examinations on an individual radiologist's basis.

Results: A total of 49 diagnostic errors were included. Top-five locations of diagnostic errors were lung (n = 8), bone (n = 8), lymph nodes (n = 5), peritoneum (n = 5), and liver (n = 4). $Workload_{normalized}$ on the days the diagnostic errors were made was on average 121% (95% confidence interval: 106% to 136%), which was significantly higher than 100% ($P = 0.008$). There was no significant upward monotonic trend in diagnostic errors over the course of the day (Mann-Kendall tau of 0.005, $P = 1.000$), and there were no other notable temporal trends either.

Conclusions: Radiologists appear to have a relative work overload when they make a diagnostic error on CT. Diagnostic errors occurred throughout the entire day, without any increase towards the end of the day.

1. Introduction

Diagnostic errors in radiology are common [1]. The retrospective error rate among radiologic examinations has been reported to be approximately 30%, with real-time errors in daily radiology practice averaging 3–5% [1]. Diagnostic errors may cause morbidity and mortality, and may lead to malpractice suits againts radiologists. Prevention of diagnostic errors is therefore paramount.

Several factors have been postulated to be associated with diagnostic errors, including workload [2]. It is important to note that the workload in radiology practice has dramatically increased and will probably keep on rising in the foreseeable future [3–6]. Given the potential association between workload and diagnostic error, there has been a discussion about instituting workload and duty limits to optimize radiologist performance and ensure patient safety [7,8]. However, experts in the field have recommended that regulating workloads without scientific principles can be more harmful than not regulating at all [7].

Although workload appears to be associated with diagnostic errors based on several previous studies [9–15], more research using real-world reading room data is necessary to support this hypothesis and to give practical guidance to radiologists and policy makers.

The purpose of this study was therefore to determine the association between workload and diagnostic errors on clinical CT scans.

2. Materials and methods

2.1. Study design and case selection

This retrospective study was approved by the local institutional review board and the requirement for informed consent was waived. All

Abbreviations: RVU, relative value unit.

* Corresponding author at: Department of Radiology, Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Hanzeplein 1, P.O. Box 30.001, 9700 RB Groningen, The Netherlands.

E-mail address: o.kasalak@umcg.nl (Ö. Kasalak).

https://doi.org/10.1016/j.ejrad.2023.111032
Received 7 August 2023; Accepted 9 August 2023
Available online 10 August 2023

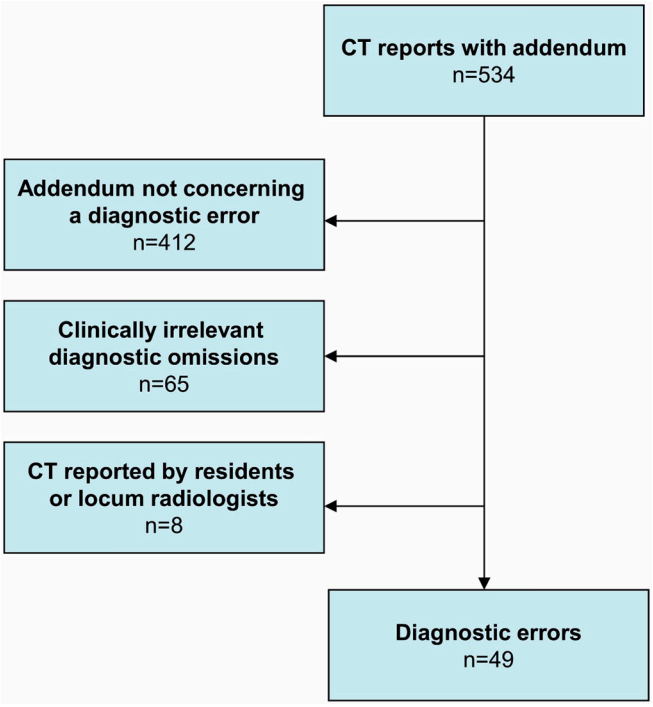

Fig. 1. Case selection flowchart.

clinical CT scans of either the brain, neck, chest, abdomen, spine, or extremities (and any combination) that were performed at a tertiary care center, reported between 7.30 a.m. and 18.00 p.m. on office days (not concerning CT scans from night or evening shifts), and that contained an addendum that was issued between 1 January 2020 and 31 March 2023, were potentially eligible for inclusion. CT scans that contained an addendum that described and corrected a diagnostic error in the original report, were included. Only perceptual errors (those in which an important abnormality was simply not seen on the images, and referred to as "diagnostic errors" in the remainder of this manuscript) were included [16]. Clinically irrelevant diagnostic omissions (e.g. non-reporting of a sebaceous cyst), as determined by a radiologist (Ö.K.) with > 5 years of post-residency clinical experience who reviewed all cases, were excluded. Cognitive errors (those in which the abnormality was visually detected but the meaning or importance of the finding was not correctly understood or appreciated) [16], cases in which the original report was signed by more than one reader, and cases that were reported by residents, fellows, or locum radiologists, were excluded.

2.2. Data collection

The relative value units (RVUs) of all examinations that were reported by the radiologist in question on the same day each diagnostic error was made (RVU_{error}), were calculated, using the 2023 RVU list of the Dutch Healthcare Authority [17]. Diffferent radiologists have different reporting capacities. To calculate each radiologist's normal reporting capacity, the average daily production of each individual radiologist ($RVU_{average}$) was calculated over an extended period of time, from 1 January 2022 to 30 June 2022. The workload on the day of each diagnostic error was then normalized for the average daily reporting capacity of each individual radiologist ($workload_{normalized}$) and calculated as: ($RVU_{error}/RVU_{average}$) × 100%. A $workload_{normalized}$ of less than 100% indicates relative work underload, while a $workload_{normalized}$ of > 100% indicates relative work overload in terms of reported examinations on an individual radiologist's basis. The time of day (categorized into 30-minute intervals) at which each diagnostic error was made was also recorded.

Table 1
Locations of 49 diagnostic errors.

Location	No.	Percentage
Lung	9	18.4%
Bone	8	16.3%
Lymph node	5	10.2%
Peritoneum	5	10.2%
Liver	4	8.2%
Adrenal gland	3	6.1%
Colon	3	6.1%
Pulmonary artery	3	6.1%
Kidney	2	4.1%
Pancreas	2	4.1%
Brain	1	2.0%
Celiac trunk	1	2.0%
Inferior vena cava	1	2.0%
Soft tissue	1	2.0%
Stomach	1	2.0%

2.3. Data analysis

A Shapiro-Wilk test was performed to confirm normal distribution of $workload_{normalized}$. It was hypothesized that diagnostic errors are more common in case of work overload, i.e. when $workload_{normalized}$ exceeds 100%. A one-sample *t*-test was performed to test for any significant differences between $workload_{normalized}$ on the days the diagnostic errors were made to a $workload_{normalized}$ of 100%. It was also hypothesized that the number of diagnostic errors would increase towards the end of the day due to fatigue. A Mann-Kendall test was performed to assess for any significant upward monotonic trend in diagnostic errors over the course of the day. *P*-values less than 0.05 were considered statistically significant. Statistical analyses were executed using MedCalc version 17.2 Software (MedCalc) and R version 4.2.3 software (R Foundation for Statistical Computing).

3. Results

3.1. Case selection

A total of 534 CT scans that were reported between 7.30 a.m. and 18.00p.m. on office days, contained an addendum that was issued between 1 January 2020 and 31 March 2023. Forty-nine diagnostic errors were finally included, as shown by the case selection flowchart in Fig. 1.

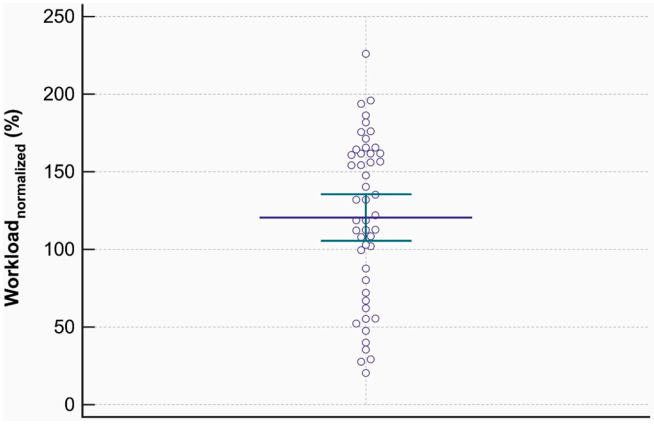

Fig. 2. Dot plot of all $workload_{normalized}$ percentages on the days the 49 diagnostic errors were made (for each individual radiologist involved). $Workload_{normalized}$ was on average 121% (indicated with the long horizontal line), with 95% confidence intervals of 106% to 136% (indicated with the short horizontal lines).

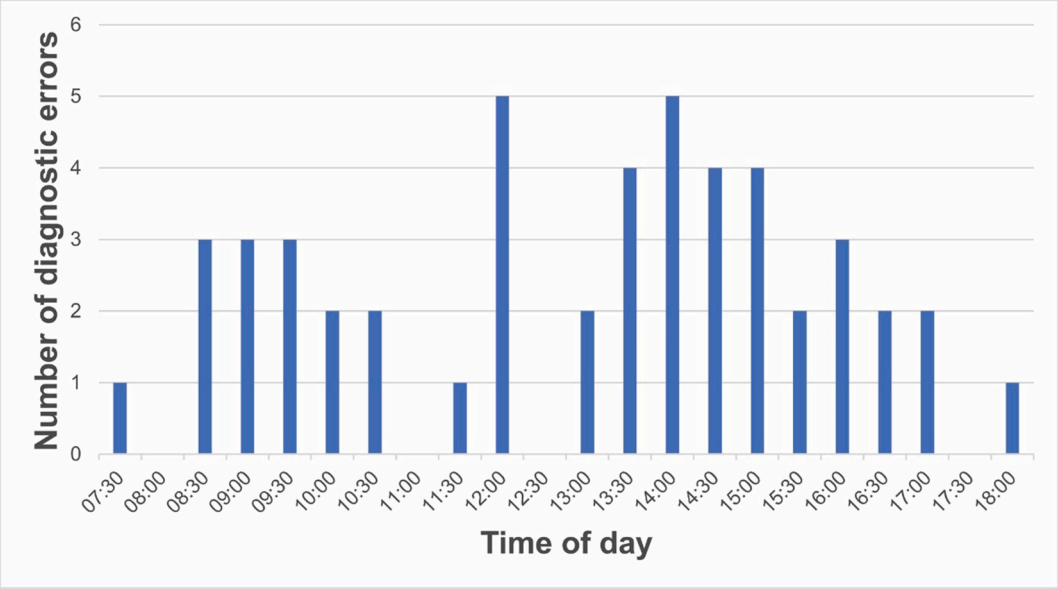

Fig. 3. Bar graph with time of day (*x*-axis) vs. number of diagnostic errors (*y*-axis).

3.2. Diagnostic errors

The 49 diagnostic errors were made by 16 different radiologists, whose main specialties were abdominal radiology (n = 5), cardiothoracic radiology (n = 4), neuroradiology (n = 3), musculoskeletal radiology (n = 2), and interventional radiology (n = 2). Top-five locations of diagnostic errors were lung (n = 8), bone (n = 8), lymph node (n = 5), peritoneum (n = 5), and liver (n = 4) (Table 1). The 49 diagnostic errors were made on CT of the chest-abdomen (n = 35), abdomen (n = 6), chest (n = 5), brain (n = 1), neck-chest-abdomen (n = 1), and abdomen-lower extremities (n = 1).

3.3. Workload

$Workload_{normalized}$ on the days the diagnostic errors were made was normally distributed ($P = 0.089$). $Workload_{normalized}$ on these days was on average 121% (95% confidence interval: 106% to 136%), which was significantly higher than 100% ($P = 0.008$). Fig. 2 shows a dot plot of all $workload_{normalized}$ percentages on the days the 49 diagnostic errors were made.

3.4. Time of day

The number of diagnostic errors per 30-minute interval ranged between 0 and 5 between 7.30 a.m. and 18.00p.m. There was no significant upward monotonic trend in diagnostic errors over the course of the day (Mann-Kendall tau of 0.005, $P = 1.000$), and there were no other notable temporal trends either, as shown in Fig. 3.

4. Discussion

The results of this study show that radiologists generally worked 21% harder (in terms of reported examinations and compared to their own average daily productivity) on days they made a diagnostic error. This indicates that radiologists should exercise caution when, for reasons like imaging backlogs, clinical pressure, and understaffing, their caseload exceeds their normal daily production. Medical imaging utilization and the need for diagnostic radiologists are expected to keep on increasing in the foreseeable future [5,6]. Health care policy makers should make efforts to ensure a healthy balance between workload and staffing, which can be regarded an important precondition for radiologists to deliver high-quality care. Nevertheless, it should be noted that workload is not the only cause of diagnostic errors, highlighted by the fact that in 15 out of 49 diagnostic errors (30.6%) in the present study, $workload_{normalized}$ was less than 100%. Interestingly, our hypothesis that the number of diagnostic errors would increase towards the end of the day due to fatigue, was not proven. This may be related to the duration of the day shift in our hospital (officially 9 h, from 8.00 a.m. to 17.00 p.m.), which could be different with more extended (>10) working hours [14].

Several previous studies have reported that reduced viewing time can lead to diagnostic error [9,10], and that both after a day of clinical reading and after an overnight shift, the diagnostic performance of radiologists seems to drop [11–13]. However, these previous studies were performed in an experimental, non-clinical setting in which radiologists were asked to evaluate a set of specific radiologic examinations that were selected for study purposes [9–13]. In another study, longer shifts and higher diagnostic examination volumes were reported to be associated with increased major interpretive discrepancies between the initial preliminary interpretation of a radiologist vs. the secondary review by (an)other radiologist(s) [14]. However, all radiologists in that study were aware that a second review would take place after their initial evaluation [14]. In yet another study, diagnostic errors were found to be associated with longer interpretation times, higher shift volumes, and weekend interpretation, but that study only applied to neuroradiologists [15]. A recent review evaluated the evidence regarding the effects of workload or duty hours on the accuracy of radiologists [7]. The authors of that review mentioned that most previous studies were primarily concerned with average group performance with respect to accuracy and interpretation times [7]. They concluded that although long shifts and so-called off-hours work may sometimes lead to poor performance and serve as a source of medical error, no research to date has provided the evidence necessary to establish appropriate limits for individual radiologists [7]. We agree that the available evidence does not support using absolute work limits in radiology practice, and that they may not work because of variations in reporting capacity between radiologists. However, using a relative workload metric that measures a radiologist's daily case load against his or her known average daily productivity, as was used in the present study, may be a potential alternative. Nevertheless, it should be noted that the average daily productivity of an individual radiologist may increase (or decrease) over time, which, in turn, will affect the calculated relative workload metric. Therefore, it may be necessary to monitor average daily productivity

numbers over more extended periods of time.

This study had several limitations. First, diagnostic errors were retrieved by means of analyzing addenda to the original CT reports. Diagnostic errors may have been missed due to underreporting or because they simply remained unnoticed. Although this probably means that not all diagnostic errors were retrieved from the time span that was investigated, it appears unlikely that any selection bias may have affected the results of our study. Second, only diagnostic errors on CT scans were included. Third, only perceptual errors were included, while cognitive errors were excluded. This approach was chosen because perceptual errors are regarded as the most consequential source of diagnostic error in radiology, as well as the most common reason for malpractice suits against radiologists [7]. Furthermore, presumed cognitive errors may be rather subjective when a gold standard for a definitive diagnosis is lacking, which is not infrequently the case. Fourth, workload was calculated based on the number of reported radiologic examinations, but there are several other factors that contribute to workload in clinical radiology practice (such as assigning protocols, checking image quality, consultations, preparing and attending multidisciplinary meetings, supervising residents, and other clinical teaching activities), which were not taken into account. Fifth, there are other factors that have been postulated to be associated with diagnostic errors [2], with some existing evidence that interruptions may affect accuracy [18], but this could not be assessed in the present study. Further research is necessary to confirm our results and to determine the interplay between potential determinants of diagnostic errors.

In conclusion, radiologists appear to have a relative work overload when they make a diagnostic error on CT. Diagnostic errors occurred throughout the entire day, without any increase towards the end of the day.

CRediT authorship contribution statement

Ömer Kasalak: Conceptualization, Writing – original draft, Writing – review & editing, Formal analysis, Visualization. **Haider Alnahwi:** Writing – original draft, Writing – review & editing, Data curation, Formal analysis. **Romy Toxopeus:** Writing – original draft, Writing – review & editing. **Derya Yakar:** Writing – original draft, Writing – review & editing. **Thomas C. Kwee:** Conceptualization, Writing – original draft, Writing – review & editing, Formal analysis, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C.S. Lee, P.G. Nagy, S.J. Weaver, D.E. Newman-Toker, Cognitive and system factors contributing to diagnostic errors in radiology, AJR Am J Roentgenol 201 (3) (2013) 611–617.
[2] J.N. Itri, R.R. Tappouni, R.O. McEachern, A.J. Pesch, S.H. Patel, Fundamentals of diagnostic error in imaging, Radiographics 38 (6) (2018) 1845–1865.
[3] C. Dan Lantsman, Y. Barash, E. Klang, L. Guranda, E. Konen, N. Tau, Trend in radiologist workload compared to number of admissions in the emergency department, Eur J Radiol 149 (2022) 110195.
[4] R.J.M. Bruls, R.M. Kwee, Workload for radiologists during on-call hours: Dramatic increase in the past 15 years, Insights Imaging 11 (2020) 121.
[5] T.C. Kwee, R.M. Kwee, Workload of diagnostic radiologists in the foreseeable future based on recent scientific advances: Growth expectations and role of artificial intelligence, Insights Imaging 12 (2021) 88.
[6] T.C. Kwee, M.T. Almaghrabi, R.M. Kwee, Diagnostic radiology and its future: What do clinicians need and think? Eur Radiol (2023) https://doi.org/10.1007/s00330-023-09897-2.
[7] R. Alexander, S. Waite, M.A. Bruno, E.A. Krupinski, L. Berlin, S. Macknik, S. Martinez-Conde, Mandating limits on workload, duty, and speed in Radiology, Radiology 304 (2) (2022) 274–282.
[8] F.J. Lexa, Duty hour limits for radiologists: It's about time, J Am Coll Radiol 18 (1) (2021) 208–210.
[9] E.E. Christensen, R.C. Murry, K. Holland, J. Reynolds, M.J. Landay, J.G. Moore, The effect of search time on perception, Radiology 138 (2) (1981) 361–365.
[10] E. Sokolovskaya, T. Shinde, R.B. Ruchman, A.J. Kwak, S. Lu, Y.K. Shariff, E. F. Wiggins, L. Talangbayan, The effect of faster reporting speed for imaging studies on the number of misses and interpretation errors: A pilot study, J Am Coll Radiol 12 (7) (2015) 683–688.
[11] E.A. Krupinski, K.S. Berbaum, R.T. Caldwell, K.M. Schartz, J. Kim, Radiology workdays reduce detection and accommodation accuracy, J Am Coll Radiol 7 (9) (2010) 698–704.
[12] E.A. Krupinski, K.S. Berbaum, R.T. Caldwell, K.M. Schartz, M.T. Madsen, D. J. Kramer, Do long radiology workdays affect nodule detection in dynamic CT interpretation? J Am Coll Radiol. 9 (3) (2012) 191–198.
[13] T.N. Hanna, M.E. Zygmont, R. Peterson, D. Theriot, H. Shekhani, J.-O. Johnson, E. A. Krupinski, The effects of fatigue from overnight shifts on radiology search patterns and diagnostic performance, J Am Coll Radiol 15 (12) (2018) 1709–1716.
[14] T.N. Hanna, C. Lamoureux, E.A. Krupinski, S. Weber, J.-O. Johnson, Effect of shift, schedule, and volume on interpretive accuracy: A retrospective analysis of 2.9 million radiologic examinations, Radiology 287 (1) (2018) 205–212.
[15] V. Ivanovic, K. Broadhead, R. Beck, Y.M. Chang, A. Paydar, G. Biddle, L. Hacein-Bey, L. Qi, Factors associated with neuroradiologic diagnostic errors at a large tertiary-care academic medical center: A case-control study, AJR Am J Roentgenol (2013) 1–8.
[16] M.A. Bruno, E.A. Walker, H.H. Abujudeh, Understanding and confronting our mistakes: The epidemiology of error in radiology and strategies for error reduction, Radiographics 35 (6) (2015) 1668–1676.
[17] Radiological Society of The Netherlands. NZa Zorgactiviteitenlijst en NVvR registratieregels. Available via: https://radiologen.nl/secties/commissie-voor-beroepsaangelegenheden/documenten/nza-zorgactiviteitenlijst-en-nvvr. Accessed on 10 July 2023.
[18] R.M. Wynn, J.L. Howe, L.C. Kelahan, A. Fong, R.W. Filice, R.M. Ratwani, The impact of interruptions on chest radiograph interpretation: Effects on reading time and accuracy, Acad Radiol 25 (12) (2018) 1515–1520.