Is work overload associated with diagnostic errors on ^{18}F-FDG-PET/CT?

Toxopeus, Romy; Kasalak, Ömer; Yakar, Derya; Noordzij, Walter; Dierckx, Rudi A J O; Kwee, Thomas C

Published in:
European Journal of Nuclear Medicine and Molecular Imaging

DOI:
10.1007/s00259-023-06543-3

European Journal of Nuclear Medicine and Molecular Imaging (2024) 51:1079–1084
https://doi.org/10.1007/s00259-023-06543-3

ORIGINAL ARTICLE

Is work overload associated with diagnostic errors on ^{18}F-FDG-PET/CT?

Romy Toxopeus[1] · Ömer Kasalak[1] · Derya Yakar[1] · Walter Noordzij[1] · Rudi A. J. O. Dierckx[1] · Thomas C. Kwee[1]

Received: 29 September 2023 / Accepted: 22 November 2023 / Published online: 30 November 2023

Abstract

Purpose To determine the association between workload and diagnostic errors on ^{18}F-FDG-PET/CT.

Materials and methods This study included 103 ^{18}F-FDG-PET/CT scans with a diagnostic error that was corrected with an addendum between March 2018 and July 2023. All scans were performed at a tertiary care center. The workload of each nuclear medicine physician or radiologist who authorized the ^{18}F-FDG-PET/CT report was determined on the day the diagnostic error was made and normalized for his or her own average daily production (workload$_{normalized}$). A workload$_{normalized}$ of more than 100% indicates that the nuclear medicine physician or radiologist had a relative work overload, while a value of less than 100% indicates a relative work underload on the day the diagnostic error was made. The time of the day the diagnostic error was made was also recorded. Workload$_{normalized}$ was compared to 100% using a signed rank sum test, with the hypothesis that it would significantly exceed 100%. A Mann–Kendall test was performed to test the hypothesis that diagnostic errors would increase over the course of the day.

Results Workload$_{normalized}$ (median of 121%, interquartile range: 71 to 146%) on the days the diagnostic errors were made was significantly higher than 100% ($P=0.014$). There was no significant upward trend in the frequency of diagnostic errors over the course of the day (Mann–Kendall tau = 0.05, $P=0.7294$).

Conclusion Work overload seems to be associated with diagnostic errors on ^{18}F-FDG-PET/CT. Diagnostic errors were encountered throughout the entire working day, without any upward trend towards the end of the day.

Keywords ^{18}F-FDG · Computed tomography · Diagnostic errors · PET-CT · Workload

Introduction

The number of diagnostic imaging examinations has increased substantially over the past decades [1, 2]. A study reported that the average radiologist interpreting CT or MRI examinations must interpret one image every 3–4 s in an 8-h workday to meet workload demands, based on data from 2010 [1]. Another more recent study on diagnostic X-ray, CT, and MRI examinations reported that the average time spent on interpreting each diagnostic image decreased from 16.0 to 2.9 s between 2005 and 2022 [2]. The workload is expected to increase further in the foreseeable future [3, 4].

Concerns have been raised that increased workload may give rise to diagnostic errors, and that regulation may be necessary to ensure patient safety [5–7]. Interestingly, a recent expert review that addressed the question if workload and duty limits should be instituted to optimize radiologist performance and patient safety, concluded that the scientific evidence needed to make meaningful rules is lacking and that regulating workloads without scientific principles can be more harmful than not regulating at all [8].

Workload in nuclear medicine departments, including ours, has also grown considerably over the past years, with PET/CT using the radiotracer ^{18}F-FDG being the most important contributor to this growth. It is currently unknown if work overload leads to diagnostic errors in imaging. Scientific evidence that supports this hypothesis may be used by stakeholders in the field to initiate or reinforce a policy in which workload and available workforce are kept in balance, rather than allowing relentless growth without adequate staffing.

✉ Thomas C. Kwee
thomaskwee@gmail.com

[1] Medical Imaging Center, Departments of Radiology and Nuclear Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

The purpose of this study was therefore to determine the association between workload and diagnostic errors on ^{18}F-FDG-PET/CT.

Materials and methods

Study design

This retrospective study was approved by the local ethical review board, and the requirement for informed consent was waived. All PET/CT reports with an addendum that were issued at a tertiary care center between March 2018 and July 2023 were potentially eligible for inclusion. The PET/CT examination had to be performed with ^{18}F-FDG as radiotracer and had to cover the region from either midthighs, midtiba, or feet to cranial vertex. Only ^{18}F-FDG-PET/CT scans with an addendum that described a perceptual error were included. Perceptual errors (henceforth simply referred to as "diagnostic errors") were defined as important abnormalities with clinical consequences, that were not seen or described in the original ^{18}F-FDG-PET/CT report [9]. Interpretive errors, defined as the incorrect interpretation of the meaning or importance of an identified abnormality [9], were excluded. ^{18}F-FDG-PET/CT scans with an addendum that addressed a dictation error or that contained additional information without any clinical consequences were also excluded. For example additional measurements, a pre-existing condition or findings which needed no further follow-up were considered as additional information with non-clinical consequences.

FDG-PET/CT acquisition and interpretation

FDG-PET/CT was performed according to European Association of Nuclear Medicine guidelines [10]. Patients fasted for at least 6 h before intravenous administration of 2–3 MBq of ^{18}F-FDG per kg body weight was administered intravenously. Blood glucose levels were checked to be less than 11 mmol/L. Three different PET/CT systems were used for image acquisition (Biograph mCT, Biograph Vision, and Biograph Vision Quadra; Siemens Healthineers). PET imaging was performed from midthighs, midtibia, or feet to cranial vertex, 60 min after ^{18}F-FDG administration. Low-dose CT (100 kV and automatic tube current modulation) was acquired for attenuation correction and anatomic imaging. Concomitant contrast-enhanced CT (100–120 kV and automatic tube current modulation) was performed in a subset of patients. Low-dose ^{18}F-FDG-PET/CT scans were interpreted by nuclear medicine physicians or radiologists certified to interpret ^{18}F-FDG-PET/CT. Concomitant diagnostic CT scans, when performed, were interpreted by radiologists or nuclear medicine physicians certified to interpret diagnostic CT.

Data extraction

Workload was quantified using relative value units (RVUs), which have been established for each imaging examination/procedure by the Dutch Healthcare Authority, in collaboration with working groups of the Dutch Society of Nuclear Medicine and Dutch Society of Radiology [11]. The RVUs of a particular imaging examination indicate the workload of that particular imaging examination relative to other imaging examinations. For example, one chest radiograph corresponds to 3.75 RVUs, one abdominal ultrasonography examination corresponds to 15 RVUs, one chest CT corresponds to 19.5 RVUs, and one low-dose ^{18}F-FDG-PET/CT corresponds to 100 RVUs, according to the 2023 guideline [11]. The workload of each individual nuclear medicine physician or radiologist who authorized the ^{18}F-FDG-PET/CT report in which the diagnostic error was made, was determined. This workload was calculated for the day the diagnostic error was made, as well as a daily average over a longer period of time (January 1st 2022 until June 30th 2022). By calculating the ratio of the workload on the day the diagnostic error was made to the workload on a daily basis over this longer period of time, a normalized workload metric ($\text{workload}_{\text{normalized}}$) was calculated. Note that this $\text{workload}_{\text{normalized}}$ metric was separately calculated for each individual nuclear medicine physician/radiologist, and not averaged over all reporting physicians. A $\text{workload}_{\text{normalized}}$ of more than 100% indicates that the nuclear medicine physician or radiologist had a relative work overload, while a value of less than 100% indicates a relative workload underload compared to his/her individual average on the day the diagnostic error was made. The time of the day the diagnostic error was made (in terms of 30-min intervals throughout the day) was also recorded for further analysis.

Data analysis

A Shapiro–Wilk test was performed to assess the (normal) distribution of $\text{workload}_{\text{normalized}}$. $\text{Workload}_{\text{normalized}}$ was then compared to 100% using either a one-sample *t*-test or a signed rank sum test, with the hypothesis that it would significantly exceed 100%. A Mann–Kendall test was performed to test the hypothesis that diagnostic errors would increase over the course of the day. Statistical significance was set at $P < 0.05$. Statistical analyses were conducted using MedCalc version 15.8 Software (MedCalc) and R version 4.3.1 software (R Foundation for Statistical Computing).

Results

Included cases

A total of 500 PET/CT reports that were issued between March 2018 and July 2023, contained an addendum. The time between the addendum and original report had a median of 5 days (interquartile range: 1 to 12.8 days). After applying the exclusion criteria, 103 ^{18}F-FDG-PET/CT scans with a diagnostic error were eventually included (Fig. 1). ^{18}F-FDG-PET/CT was performed from midthighs to cranial vertex in 87 (84.5%) cases, from feet to cranial vertex in 14 (13.6%) cases, and from midtibia to cranial vertex in 2 (1.9%) cases. A concomitant diagnostic CT was performed in 44 (42.7%) cases. Most ^{18}F-FDG-PET/CT scans were performed because of an oncologic indication ($n = 75$; 72.8%), followed by infection ($n = 20$; 19.4%), inflammation ($n = 7$; 6.8%) and others ($n = 1$; 1.0%). Note that a total of 26,157 clinical ^{18}F-FDG-PET/CT scans (with an anatomic coverage from at least midthighs to cranial vertex) were performed and reported in our hospital in this same period.

Diagnostic errors

The diagnostic errors were made by 9 different radiologists, 8 different nuclear medicine physicians and 1 nuclear radiologist (i.e., a radiologist with completed nuclear medicine specialty training in the Dutch integrated nuclear medicine and radiology residency training program). Their postresidency experience ranged from 0 to 36 years, with a median of 5 years. The locations of the diagnostic errors are shown in Table 1, with the top-three locations concerning the bone ($n = 17$, 15.9%), lung ($n = 12$, 11.2%) and head-neck region ($n = 11$, 10.3%). Note that 4 ^{18}F-FDG-PET/CT scans contained 2 diagnostic errors.

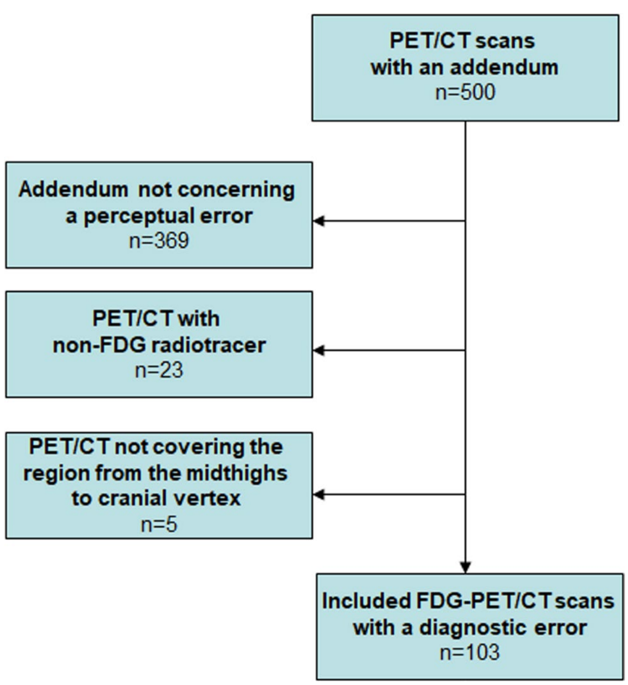

Fig. 1 Case selection flow chart

Workload

The total number of examinations that were reported by each nuclear medicine physician/radiologist in the reference period (January 1st 2022 until June 30th 2022) that were used to calculate $\text{workload}_{\text{normalized}}$ are shown in Table 2. The average daily workload per nuclear medicine physician/radiologist in this reference period was 747 ± 380 RVUs (range: 59 to 1705 RVUs). $\text{Workload}_{\text{normalized}}$ did not have a normal distribution ($P = 0.032$). $\text{Workload}_{\text{normalized}}$ (median of 121%, interquartile range: 71% to 146%) on the days the diagnostic errors were made was significantly higher than 100% ($P = 0.014$). Figure 2 shows a Box-and-whisker plot of the $\text{workload}_{\text{normalized}}$ percentages on the days the 103 diagnostic errors were made.

Table 1 Locations of diagnostic errors on 103 ^{18}F-FDG-PET/CT scans

Location of diagnostic error	No	% of total
Bone	17	15.9
Lung	12	11.2
Head-neck region	11	10.3
Bowel (small intestine/colon/rectum)	10	9.3
Heart/blood vessels	10	9.3
Kidney	5	4.7
Lymph node	5	4.7
Soft tissue	5	4.7
Liver	4	3.7
Prostate	4	3.7
Stomach	4	3.7
Brain	3	2.8
Ovary	3	2.8
Spleen	3	2.8
Adrenal gland	2	1.9
Breast	2	1.9
Muscle	2	1.9
Thyroid	2	1.9
Mesentery	1	0.9
Urinary Bladder	1	0.9
Uterus	1	0.9

Table 2 Overview of the total numbers of examinations that were reported by each nuclear medicine physician/radiologist in the reference period (January 1st 2022 until June 30th 2022) that was used to calculate workload$_{normalized}$. Subsequently, for each individual radiologist, all reported examination was multiplied by their corresponding RVUs, and the total sum was then divided by the number of days between January 1st 2022, and June 30th 2022, to obtain the average daily workload per nuclear medicine physician/radiologist in this reference period. Finally, by calculating the ratio of the workload on the day the diagnostic error was made to this average daily workload, a normalized workload metric (workload$_{normalized}$) was obtained, for each individual nuclear medicine physician/radiologist

Nuclear medicine physician/radiologist	Conventional radiography	Ultrasonography	CT	MRI	Conventional scintigraphy	PET	Total
A	133	0	58	0	75	255	521
B	1962	4	642	0	0	11	2619
C	122	0	6	0	77	197	402
D	836	567	885	452	0	14	2754
E	13	0	4	0	9	23	49
F	2985	101	1100	77	0	48	4311
G	1627	350	535	313	0	32	2857
H	115	0	154	0	118	364	751
I	316	366	830	224	0	97	1833
J	146	0	8	0	90	203	447
K	206	0	13	0	201	270	690
L	71	0	12	0	304	205	592
M	51	131	585	759	0	7	1533
N	213	0	132	0	247	459	1051
O	18	0	5	0	69	142	234
P	632	285	600	9	0	1228	2754
Q	40	158	491	784	0	9	1482
R	226	207	251	216	0	7	907

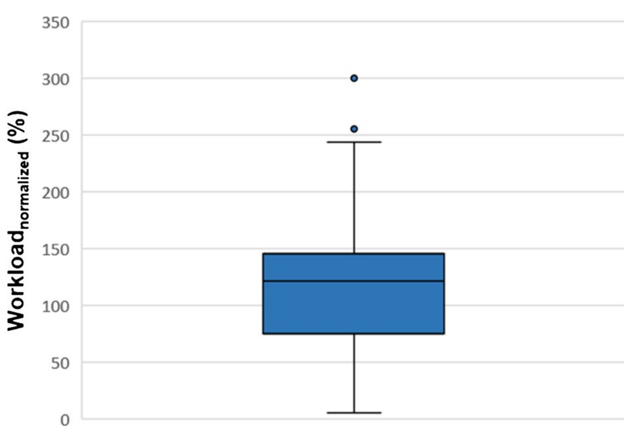

Fig. 2 Box-and-whisker plot of the workload$_{normalized}$ percentages on the days the 103 diagnostic errors were made on ^{18}F-FDG-PET/CT, with median (middle line of box), quartiles (top and bottom lines of box), upper extreme value (upper whisker), lower extreme value (lower whisker), and outliers (circles)

Temporal tends

The time of the day at which diagnostic errors were made ranged from 08:00 until 19:30, and the number of diagnostic errors in 30-min intervals ranged between 0 and 10 (Fig. 3). There was no significant upward trend in the frequency of diagnostic errors over the course of the day (Mann–Kendall tau = 0.05, $P = 0.7294$).

Discussion

The results of this study show that nuclear medicine physicians and radiologists have a relative work overload (on average approximately 21%) on the day they made a diagnostic error, when compared to their average daily production over a longer period of time. This finding adds evidence to the notion that over-intensification of work may provoke diagnostic errors. However, diagnostic errors did not show any particular distribution pattern over the course of the day, and our hypothesis that diagnostic errors would increase at the end of the day due to fatigue was not proven. Nevertheless, visual inspection of Fig. 3 suggests that most diagnostic errors were made between 12.00 and 16.30. This can be explained by the fact that most ^{18}F-FDG-PET/CT scans were reconstructed and ready for reading and reporting in this time period. In addition, the majority of the larger multidisciplinary team meetings (MDTMs) are scheduled around lunch time. Upon return from the MDTM, the list of scans to be reported is often very long, which may stress the physicians to finalize the reports quickly. Workload, and in

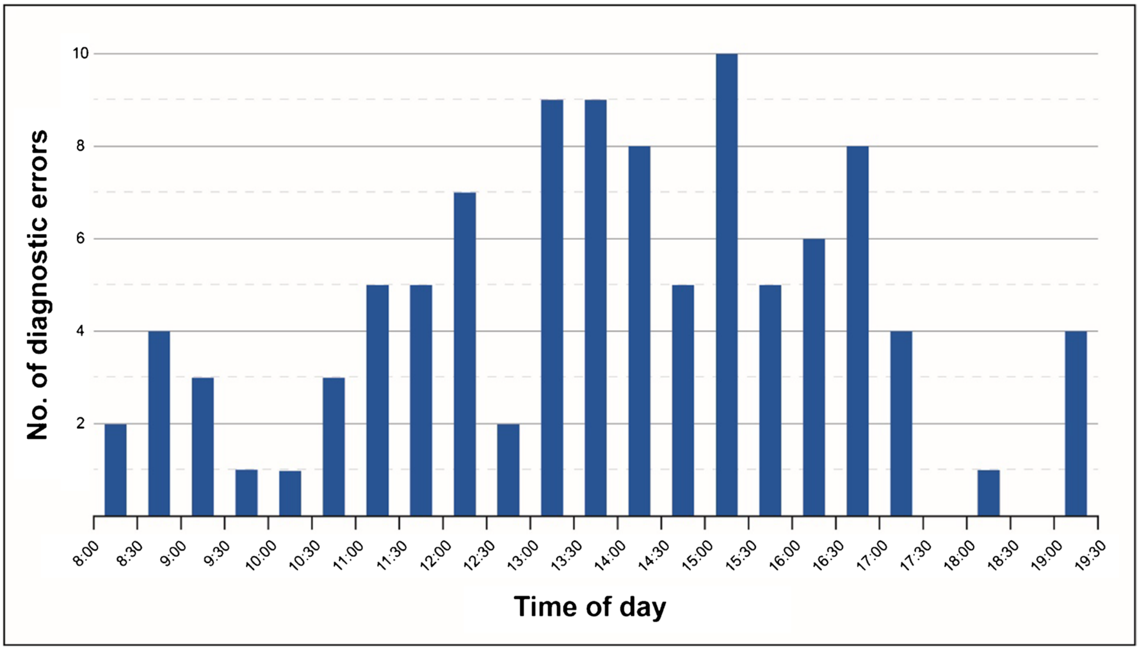

Fig. 3 Bar graph showing the number of diagnostic errors over the course of the day (according to 30-min intervals)

particular requests for PET/CT, have kept on increasing over the past years, both in our hospital and undoubtedly in many other centers worldwide. Our results suggest that a proportional increase in staffing may be required to ensure the quality and safety of ^{18}F-FDG-PET/CT reporting. Notably, the number of physicians available for reporting ^{18}F-FDG-PET/CT scans in our department did not increase between 2018 and 2023. Artificial intelligence is regarded as a potential means to improve efficiency and safety in healthcare, but whether or not it may be useful to nuclear medicine physicians and radiologists to cope with their increasing workload in the foreseeable future, remains unclear.

The number of studies using real world clinical practice data to determine the association between workload and diagnostic errors in medical imaging is limited. In a study by Ivanovic et al. [12], 654 neuroradiological CT or MRI reports with a diagnostic error were reviewed. The mean number of interpreted studies during shifts when a diagnostic error was made (46.6) was significantly higher ($P<0.001$) than during shifts when no diagnostic error was made (34.1) [12]. In another study by Ivanovic et al. [13] that included 654 neuroradiological CT or MRI examinations with a diagnostic error and 1,019 without, it was found that a diagnostic error was independently associated with interpretation time (odds ratio [OR] of 1.18, $P=0.003$), shift volume (OR of 1.27, $P<0.001$), and weekend interpretation (OR of 1.69, $P=0.02$). In a sub-analysis, diagnostic errors showed independent associations on weekdays with interpretation time (OR of 1.18, $P=0.003$) and shift volume (OR of 1.27, $P<0.001$), whereas such associations were not observed on weekends [13]. In a study by Kasalak et al. [14] that used the same methodology as ours, the relationship between workload and diagnostic errors on clinical (body) CT scans was investigated. In that study with 49 diagnostic CT errors, workload$_{normalized}$ on the days the diagnostic errors were made was on average 121%, and significantly higher ($P=0.008$) than 100%. Furthermore, there was no significant upward monotonic trend in diagnostic CT 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 [14]. The results by Ivanovic et al. [12, 13] and Kasalak et al. [14] resonate with those of the present study. However, the topics of work overload and diagnostic errors have remained relatively underexposed in the nuclear medicine community so far.

The present study had some limitations. First, diagnostic errors on ^{18}F-FDG-PET/CT scans were identified by screening addenda to the original reports. The number of diagnostic errors that was identified in this way was probably an underestimations of the total number of diagnostic errors on ^{18}F-FDG-PET/CT in our study period. Nevertheless, this is unlikely to have affected our results. Second, this study was performed at a tertiary care center with the far majority of ^{18}F-FDG-PET/CT scans performed due to an oncologic indication. The results may be different in other patient populations. Third, the association between workload and diagnostic errors was assessed without correction for other potentially relevant determinants. Further research is required to confirm our results and to investigate if workload remains as an independent determinant of diagnostic errors when also considering other factors such as the experience of the readers, the quality of the ^{18}F-FDG-PET/CT requests, and work interruptions. Fourth, workload was calculated based on the number and complexity of reported imaging examinations, but there are several other factors that contribute to workload in clinical practice. The other activities include protocolling imaging studies, communication with technicians and other physicians, and preparing and attending MDTMs. Teaching and training also contribute to workload in many hospitals. However, their contribution to the total workload could not be assessed and taken into account. Therefore, our results may be an underestimation of the true workload and its impact on diagnostic errors. Fifth, it remains unclear if the use of newer generations of PET/CT systems (which allow for faster scanning and/or provide better image quality) was associated with less or more diagnostic errors, because the relatively small sample sizes did not allow for any meaningful statistical analysis.

In conclusion, work overload seems to be associated with diagnostic errors on ^{18}F-FDG-PET/CT. Diagnostic errors were encountered throughout the entire working day, without any upward trend towards the end of the day.

Declarations

Competing Interests The authors have no conflicts of interested related to this article.

References

1. McDonald RJ, Schwartz KM, Eckel LJ, Diehn FE, Hunt CH, Bartholmai BJ, Erickson BJ, Kallmes DF. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. Acad Radiol. 2015;22:1191–8.
2. Peng YC, Lee WJ, Chang YC, Chan WP, Chen SJ. Radiologist burnout: Trends in medical imaging utilization under the national health insurance system with the universal code bundling strategy in an academic tertiary medical centre. Eur J Radiol. 2022;157:110596.
3. Kwee TC, Kwee RM. Workload of diagnostic radiologists in the foreseeable future based on recent scientific advances: growth expectations and role of artificial intelligence. Insights Imaging. 2021;12:88.
4. Kwee TC, Almaghrabi MT, Kwee RM. Diagnostic radiology and its future: what do clinicians need and think? Eur Radiol. 2023. https://doi.org/10.1007/s00330-023-09897-2.
5. Lee CS, Nagy PG, Weaver SJ, Newman-Toker DE. Cognitive and system factors contributing to diagnostic errors in radiology. AJR Am J Roentgenol. 2013;201(3):611–7.
6. Taylor-Phillips S, Stinton C. Fatigue in radiology: a fertile area for future research. Br J Radiol. 2019;92:20190043.
7. Saliba T, Simoni P, Boitsios G. Commentary: How much further can radiologists be pushed? Pediatr Radiol. 2023. https://doi.org/10.1007/s00247-023-05741-3.
8. Alexander R, Waite S, Bruno MA, Krupinski EA, Berlin L, Macknik S, Martinez-Conde S. Mandating limits on workload, duty, and speed in radiology. Radiology. 2022;304:274–82.
9. Itri JN, Patel SH. Heuristics and cognitive error in medical imaging. AJR Am J Roentgenol. 2018;210:1097–105.
10. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, Verzijlbergen FJ, Barrington SF, Pike LC, Weber WA, Stroobants S, Delbeke D, Donohoe KJ, Holbrook S, Graham MM, Testanera G, Hoekstra OS, Zijlstra J, Visser E, Hoekstra CJ, Pruim J, Willemsen A, Arends B, Kotzerke J, Bockisch A, Beyer T, Chiti A, Krause BJ. European Association of Nuclear Medicine (EANM). FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42:328–54.
11. Radiological Society of The Netherlands. Nza Zorgactiviteitenlijst en NVvR registratieregels. Available via: https://radiologen.nl/secties/commissie-voorberoepsaangelegenheden/documenten/nza-zorgactiviteitenlijst-en-nvvr. Accessed on 10 August 2023.
12. Ivanovic V, Paydar A, Chang YM, Broadhead K, Smullen D, Klein A, Hacein-Bey L. Impact of shift volume on neuroradiology diagnostic errors at a large tertiary academic center. Acad Radiol. 2023;30:1584–8.
13. Ivanovic V, Broadhead K, Beck R, Chang YM, Paydar A, Biddle G, Hacein-Bey L, Qi L. Factors associated with neuroradiologic diagnostic errors at a large tertiary-care academic medical center: a case-control Study. AJR Am J Roentgenol. 2023;221:355–62.
14. Kasalak Ö, Alnahwi H, Toxopeus R, Pennings JP, Yakar D, Kwee TC. Work overload and diagnostic errors in radiology. Eur J Radiol. 2023;167: 111032.