BMJ Open

Association of volume of self-directed versus assigned interpretive work with diagnostic performance of radiologists: an observational study

Shiori Amemiya 🟢 , Harushi Mori, Hidemasa Takao, Osamu Abe

Radiology, The University of Tokyo, Tokyo, Japan

Correspondence to
Dr Shiori Amemiya;
amemiya-tky@umin.ac.jp

ABSTRACT

Objectives To understand the sources of variability in diagnostic performance among experienced radiologists.
Design All prostate MRI examinations performed between 2016 and 2018 were retrospectively reviewed.
Setting University hospital in Japan.
Participants Data derived from 334 pathology-proven cases (male, mean age: 70 years; range: 35–90 years) that were interpreted by 10 experienced radiologists were subjected to the analysis.
Primary and secondary outcome measures Diagnostic performance measures of the radiologists were compared with candidate factors, including interpretive volume of prostate MRIs, volume of self-directed and assigned total annual interpretive work, and years of experience. The potential influence of fatigue was also evaluated by examining the effect of the report's issue time.
Results There were 186 prostate cancer cases. Performance was based on accuracy, sensitivity and specificity (86%, 85% and 84%, respectively). While performance was not correlated with the volume of prostate MRIs, per se (ρ=−0.15, p=0.69; ρ=−0.01, p=0.99; ρ=−0.33, p=0.36) or the total MRIs assigned for each radiologist (p>0.6) or years of experience (p>0.4), all measures were strongly correlated with voluntary work represented by the interpretive volume of abdominal CTs (r=0.79, p<0.01; r=0.80, p<0.01; r=0.64, p=0.048). The performance did not differ based on the issue time of the report (morning, afternoon and evening) (χ^2(2)=3.65, p=0.16).
Conclusions Greater autonomy, represented as enhanced self-directed interpretive work, was most significantly correlated with the performance of prostate MRI interpretation. The lack of a correlation between the performance and assigned volume confirms the complexity of human learning. Together, these findings support the hypothesis that successful promotion of internal drivers could have a pervasive positive impact on improving diagnostic performance.

INTRODUCTION

Accurate diagnosis is central to appropriate and effective patient care,[1 2] and medical diagnosis based on imaging examinations in radiology often plays a critical role to achieve this goal. However, despite advances in technology and various strategies proposed to

---

Strengths and limitations of this study

► This study examines the factors associated with the diagnostic performance of experienced radiologists using prostate MRI examinations with a pathological confirmation.
► In addition to the interpretive volume of prostate MRIs, years of experience of the radiologists and the influence of fatigue, the motivation for interpretive work that was objectively quantified as the volume of self-directed CT interpretation was also assessed as a candidate factor.
► The limitations concern the case study research design focusing on the radiologists in a single institution in Japan.

---

overcome the problem, the rate of errors have changed little over the last 50 years,[3–5] which is estimated to be as high as 30%.[3 4 6 7]

Traditionally, most efforts have focused on intensive education and continuous training of radiologists,[3] For example, as an attempt to improve the performance of mammography, particularly sensitivity, many countries have adopted minimum annual interpretive volume requirements for physicians. While the hypothesis suggesting a volume–expertise relationship is supported by some studies,[8 9] most studies retrospectively examining the actual relationship showed that volume did not explain much of the observed inter-radiologist performance variability. Although higher-volume readers had lower false-positive rates, no sensitivity difference was found,[10–12] indicating the need to consider the learning characteristics or the nature of human errors.

The intricacy of the human errors in healthcare has long been emphasised.[13 14] As for diagnostic errors, a recent report by the National Academy of Medicine (then called the Institute of Medicine) articulates the need for a more in-depth measure, such as

considering advances in the learning sciences.[1 15] However, the problem has not been fully appreciated in clinical practice, perhaps, partly due to the lack of supporting data.[1]

The present study aimed to understand the sources of variability in diagnostic performance among experienced radiologists. Towards that end, we retrospectively reviewed the diagnostic performance for prostate cancer detection using MRI, which had the highest rate of pathological confirmation among a range of imaging examinations at our institution. In addition to the annual interpretive volume of prostate MRIs, we examined the volume of self-directed and assigned total annual interpretive work of the radiologists as possible factors characterising performance variability. The potential influence of fatigue was also evaluated by analysing the time each report was issued.

METHODS
Overview
The data included prostate MRIs performed between January 2016 and December 2018 at our university hospital (substantially all prostate MRIs ever performed) that were interpreted by 10 board-certified radiologists working at the hospital as full-time employees paid on a salary basis. Exclusion criteria were any prior biopsy or surgical intervention of the prostate and severe artefacts in images, which excluded nine and two cases, respectively. For each weekday morning and afternoon, one of the radiologists was in charge of all MRI examinations performed and issued the final reports of the studies. Prostate examinations were preferentially assigned to a slot of an abdominal radiologist, rather than a neuroradiologist.

All prostate MRIs were performed using 3T scanners with T2-weighted images (T2WI), diffusion-weighted images (b=0/1500 and apparent diffusion coefficient map calculated from the two), with dynamic contrast-enhanced (DCE) T1-weighted images in 247. All radiologists had access to the electronic health record that contains all examinations results, including prostate-specific antigen. Among the patients who had a prostate MRI, those who were considered at a higher risk of having prostate cancer and who hoped underwent an 18-core systematic biopsy combined with an MRI-US fusion-targeted biopsy. If indicated, biopsy-positive cases further underwent total prostatectomy. The surgical specimens were examined by the pathologists who delineated the cancer margin on a macroscopic picture of each specimen, which enabled the radiological–pathological correlation. The study was approved by the institutional review boards as a retrospective study, for which informed consent was waived.

Diagnostic performance
Diagnostic performance was based on the radiologist's assessment for biopsy proven cases (positive, negative or inconclusive, which corresponds to Prostate Imaging Reporting and Data System (PI-RADS) V.2[16] category of 4 and 5; 1 and 2; and 3, respectively) and determined by applying the following rules in this order: (1) if indicating at least one cancer lesion, classified as true positive (TP), irrespective of confidence level, that is, category 3 lesion was also considered as TP as long as it was pointed out as a possible cancer lesion, (2) if not indicating any definite cancer in a cancer-free prostate, classified as true negative (TN), (3) if failing to detect any cancer lesions, classified as false negative (FN), and (4) if erroneously indicating a cancer lesion, classified as false positive (FP). To compensate for MRI-undetectable cases, all FP MRIs were reviewed by two radiologists who were blinded to the results. If both two failed to detect all the lesions, the case was counted as TN. Based on these parameters, accuracy = (TP +TN)/(FP +FN), sensitivity and specificity for each radiologist were measured.

Candidate factors
Data on the annual interpretive volume of prostate MRIs, as well as that of total diagnostic MRI and CT examinations during the same period, were collected. Interpretation of CT examinations was also a part of the radiologists' duties. For each weekday, five to six radiologists were in charge of CT interpretations and were obliged to read all CTs, but with no assignment, unlike in the case of MRI, nor any individual quotas nor performance incentive. Therefore, the interpretive volume of CTs represents the volume of self-directed diagnostic work, in contrast to that of MRIs, which are semi-automatically allotted to each radiologist.

Other factors were considered, including the self-reported estimated number of a lifetime interpretive volume of prostate MRI at any hospital, subspecialty and years of experience as a radiologist. Because the rate of inconclusive diagnosis, irrespective of the reasons, directly affects the diagnostic performance, it was also examined as a possible factor. The effect of the rate of positive exams was also evaluated. The issue time of the reports was collected to explore the potential impact of fatigue. Since the starting time of the work is 08:30 for all, the issue time of report can generally reflect the cumulative working hours on the day.

In addition to the objective parameters, subjective predictive scores for each interpreter were obtained by asking the 25 colleagues (radiologists), who had worked with the interpreters for at least 1 year at the same hospital, to anonymously suggest three or five of the best interpreters of prostate MRIs for cancer detection, and by summing up the numbers a radiologist was nominated as one of the best interpreters.

Statistical analysis
For each performance measure, we calculated Pearson's correlation coefficients, r or Spearman's correlation coefficient, ρ depending on the parameter's distribution to investigate the relationship between the diagnostic performance and possible factors. Backward stepwise regression analysis based on Akaike information criterion was also used to select the explanatory variables with statistically significant effects on each performance measure from among the
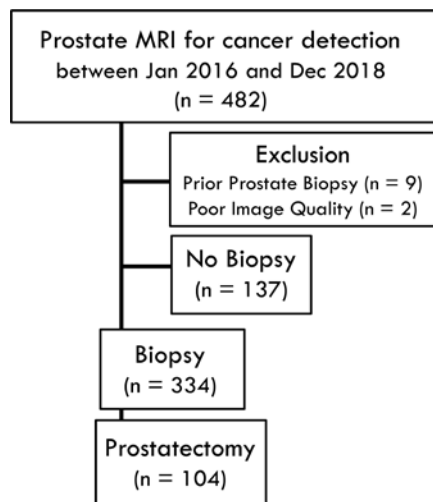
Figure 1 Flowchart of the study population.

possible confounding factors, including the rate of inconclusive diagnosis, the interpretive volume of prostate MRIs, all MRIs, or CTs, and the radiologists' years of experience. The between-group difference of diagnostic performance based on subspecialty was also tested. χ^2 test was used to assess the effect of issue time of each report classified as morning (08:00–13:00), afternoon (13:00–17:00) and evening (17:00–24:00). The diagnostic performance was also compared between the examinations performed with and without contrast-enhanced T1-weighted images using a χ^2 test. All data were analysed using IBM SPSS V.22.0 software. A two-sided p value <0.05 was considered to indicate statistical significance for all analyses.

Patient and public involvement

There were no participants involved in the development of this study.

RESULTS

The study data and profiles of the radiologists

The study included 471 consecutive MRI examinations obtained from 471 men (mean: 70 years; range: 35–90 years). Three hundred thirty-four cases underwent a biopsy within 6 months; primary prostate cancer was detected in 186 cases, which was followed by a total prostatectomy in 104 (figure 1).

Ten radiologists (nine men, mean: 40 years; range: 36–46 years) who had interpreted prostate MRIs at least for 1 year during the time period, were included in the analysis. All had worked as a diagnostic radiologist for at least 10 years (mean: 14 years; range: 10–21 years), mostly at an academic institution rather than at a community-based hospital. Six of the radiologists specialised in abdominal imaging; the remainder were neuroradiologists; none specialised in genitourinary radiology. The average annual interpretive volume of MRI examinations assigned for each radiologist was 2120 (range: 925–3381), including 16 prostate MRIs (range: 4–47), while the volume of the abdominal CT examinations self-directedly interpreted by the radiologists was 3300 (range: 1376–4906). Self-reported lifetime

interpretive volume of prostate MRIs in other hospitals was about 50–100 in nine and 250 in one.

Prostate MRI interpretation performance and candidate factors

Overall diagnostic performance of the prostate MRI interpretation corrected for MR-negative cases, measured as accuracy, sensitivity and specificity were 86%, 85% and 84% (TP/TN/FP/FN/MR-negative: 131/125/23/24/31), respectively. Uncorrected accuracy and sensitivity were 77% and 70%, respectively. The rate of positive examinations was relatively homogenous (49%±10%), and not significantly correlated with the performance (p>0.5). No significant difference of diagnostic performance was found between the examinations with and without DCE imaging ($\chi^2(2)=0.94$, p=0.625).

In terms of the radiologists' characteristics, the number of years of experience was not correlated with the diagnostic performance (p>0.4), nor was there a significant group difference based on subspecialty (t=−0.12, p=0.91; t=0.32, p=0.76; t=−0.51, p=0.62). While the diagnostic performance was not significantly correlated with the interpretive volume of prostate MRIs, per se (ρ=−0.15, p=0.69; ρ=−0.01, p=0.99; ρ=−0.33, p=0.36) (figure 2) or that of lifetime prostate MRIs (ρ=−0.41, p=0.24; ρ=−0.41, p=0.24; ρ=−0.41, p=0.24) or total MRIs (r=0.02, p=0.95; r=0.14, p=0.70; r=−0.12, p=0.75), all the performance measures were positively correlated with the annual interpretive volume of abdominal CTs (r=0.79, p<0.01; r=0.80, p<0.01; r=0.64, p=0.048) (figure 3). The same trend was replicated, even when the accuracy or sensitivity was not corrected for the MR-negative cases.

The rate of inconclusive diagnosis was 23%±8%, which showed a weak negative correlation with the diagnostic performance (r=−0.23, p=0.52; r=−0.13, p=0.73; r=−0.23; p=0.53), suggesting that it is unlikely that inconclusive diagnosis led to high performance. Rather, it was significantly negatively correlated with the years of experience (r=−0.72, p=0.02) or the volume of total MRIs (r=−0.66, p=0.04). It also had a weak negative correlation with the interpretive volume of prostate MRIs (ρ=−0.62, p=0.06).

The predictive performance scores given by suggesting the top three and top five interpreters were 7.4±7.9 and 11.3±8.1, respectively; the scores were not correlated with any of the actual performance measures (top 3: ρ=−0.30, p=0.40; ρ=0.01, p=0.99; ρ=0.53, p=0.12; top 5: r=0.24, p=0.50; r=−0.03, p=0.93; r=0.45, p=0.19). Stepwise regression analysis only indicated the volume of CTs as a significant factor for all the performance measures.

The rate of correct diagnosis based on the issue time of each report (morning/afternoon/evening: 91%/82%/89%) did not significantly differ to each other ($\chi^2(2)=3.65$, p=0.16) (figure 4).

DISCUSSIONS

In the present study, the accuracy of the radiologists interpreting prostate MRIs was 86%, which is equivalent or better
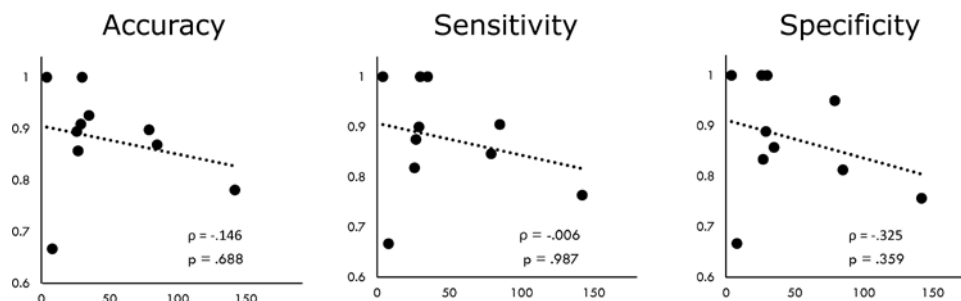
Figure 2 Diagnostic performance measures of prostate MRIs according to the interpretive volume of the examination, per se. There was no significant correlation between the performance and interpretive volume of prostate MRIs. Spearman's ρ and p values were as follows: accuracy, ρ=–0.15, p=0.69; sensitivity: ρ=–0.01, p=0.99; specificity: ρ=–033, p=0.36.

compared with the results of recent studies on the inconsistency or errors of diagnosis among experienced radiologists.[17 18] The diagnostic performance was not significantly correlated with the interpretive volume of the prostate MRIs, per se. Although the volume showed a weak negative correlation with the rate of inconclusive performance, possibly suggesting the 'practice makes confidence' effect, it did not necessarily lead to improved performance. The finding is partly consistent with previous studies that failed to confirm a direct relationship between the diagnostic volume and the sensitivity.[10–12] However, given the fact that the interpretive volume of prostate MRI was generally small and that none specialised in genitourinary radiology, the weak negative correlation might at least partly reflect overconfidence due to lack of enough experience with proper feedback that could have helped radiologists stay alert even after getting used to the prostate examinations. The fact that the actual performance was considerably different from our predictions also gave us an impression that the performance is less likely to reflect the potential of the radiologists.

To our surprise, the volume of CTs interpreted during the same period of time was most significantly associated with the performance. On the other hand, neither the interpretive volume of all MRIs nor the years of experience were significant factors. Generally speaking, it is difficult to postulate the direct effect of reading CT examinations on improving the diagnostic performance of prostate MRIs. This is because the basic knowledge required to interpret these two types of examinations is not well overlapped. The lack of a significant correlation

between the volume of total MRIs and the interpretive performance of prostate MRIs also questions the simple assumption of non-specific reading effect. Therefore, it is more reasonable to consider the relationship to be a spurious correlation caused by a third factor that affects both the prostate MRI interpretation performance and the interpretive volume of CTs.

As a vital clue to revealing the hidden factor, a distinctive characteristic of the CT interpretation is that although it was a part of the duties, there was no quota nor any type of incentive for reading more examinations. Therefore, the radiologists had to be personally motivated to keep reading. From a behavioural scientific point of view, it is not difficult to understand why such behaviour that is more dependent on internalised motivation rather than on an external regulation, namely, assignment of duty, was more strongly associated with higher performance. More concretely, it is the core concept of the self-determination theory, one of the most widely accepted theories in contemporary behavioural science.[19]

As an approach to human motivation, self-determination theory highlights the importance of our basic organismic needs as the drivers for intrinsic motivation. These include: (1) autonomy or self-determination, which refers to being self-initiating and self-regulating of one's actions, (2) competence (ie, self-efficacy and mastery), and (3) relatedness, which involves developing secure and satisfying connections with others in one's social milieu.[20] All these factors are grounded in the evolutionary benefits in terms of survival. According to self-determination theory, internalised motivation associated with greater
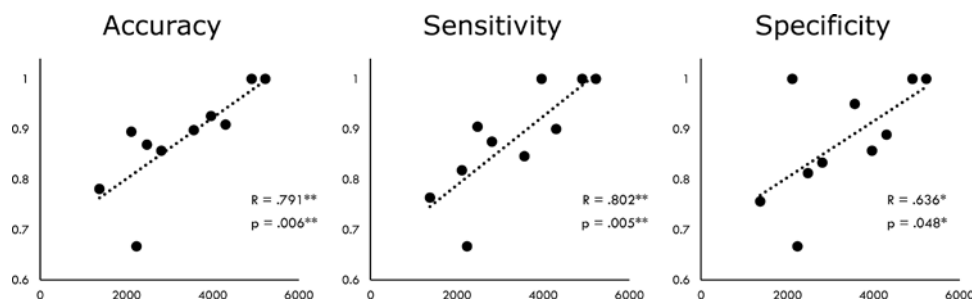

Figure 3 Diagnostic performance measures of prostate MRIs according to the annual interpretive volume of abdominal CTs. All measures showed significant correlation with the interpretive volume of CTs. Pearson's r and p values were as follows: accuracy, r=0.79, p<0.01; sensitivity, r=0.80, p<0.01; specificity, r=0.64, p=0.048.
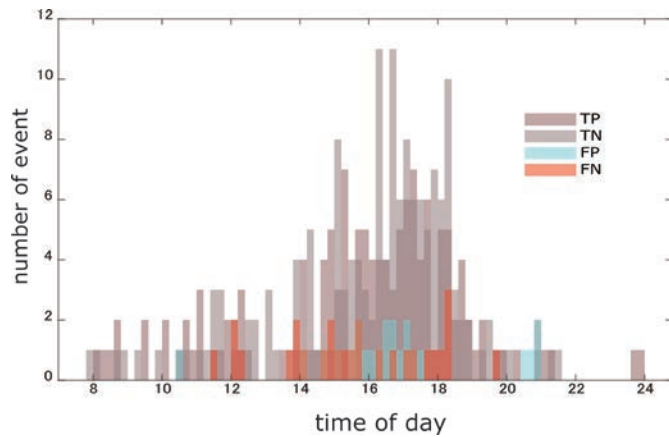
Figure 4 Diagnostic performance of prostate MRIs according to the issue time of the reports. The rate of correct diagnosis based on the issue time of each report was: 91% for morning, 82% for afternoon and 89% for evening; the differences were not statistically significant ($\chi^2(2)$=3.65, p=0.16). FN,false negative; FP, false positive; TN, true negative; TP, true positive.

autonomy leads to greater persistence, more positive self-perceptions and better quality of engagement.[21] Empirically, the theory has been proven to be applicable in diverse domains, including school education, businesses and healthcare.[22–24] Conversely, it is also well known that external rewards, especially monetary incentives, could undermine intrinsic motivation.[25] Although it is still controversial, the undermining effect is considered as one of the main reasons why the effectiveness of financial incentives in clinical medicine is only supported by modest and inconsistent evidence.[26–30]

On considering the risks and possible benefits associated with interpreting more examinations, we deem it appropriate to regard the difference to be the relative degree of success in the internalisation of external motivation for diagnostic work. The fact that most of our activities are, strictly speaking, not considered to be intrinsically motivated[21] leaves room for intervention. Then, what could lead us to motivate ourselves and how? Although not through an experiment, a brief debriefing revealed that desire for competence was not an essential factor for those who interpreted more CTs. This might not be surprising given that the radiologists were experienced experts rather than trainees. Instead, they reported that they were motivated by a sense of responsibility or relatedness, that is, someone needs to read the examination since the patient or attending physician is waiting for the report. They also reported that such a notion was reinforced via communications with physicians providing information or giving feedback, as well as with their colleagues within the department. These findings agree with the recent recommendations from the Institute of Medicine that highlight the potential benefit of teamwork for improving clinical diagnosis, and support the development of an organisational culture that values open discussion and feedback to leverage the intrinsic motivation of medical professionals.[1 15]

Healthcare systems vary from country to country. The actual status of clinical practice involving physicians also varies according to social structure, culture and customs. Our data acquired from a small number of radiologists in a single institute might not necessarily reflect the representative behaviour or views of the radiologists or physicians in general. The rate of subspecialist radiologists varies a lot. While it is common in many European countries including the UK to practice as a general or multispecialty radiologist,[31 32] the rate of subspecialists is much higher in the USA.[31 33] As for prostate MRI, the cumulative number of lifetime interpretations was generally small in the present study, although it exceeded the approximate number required to reach a plateau of the learning curve.[34] The status might be substantially different from the facilities where prostate MRIs are exclusively interpreted by genitourinary radiologists. Regarding CT interpretation, we have long adopted a highly flexible duty system that does not assign individual quotas, to accommodate for our variable and multitasking working condition at an academic institution, as well as to avoid possible negative effects caused by the pursuit of volume. Such a system would also be exceedingly rare. However, given the generality of the problems concerning human behaviour, we assume that the generalisability of the results might not rely much on the representability of our clinical setting.

There are some other limitations to the present study. Regarding the diagnostic performance, the degree of difficulty differs for each case. Nevertheless, for the sensitivity, although we compensated for MR-negative cases, which should have at least partly controlled the varying degree of difficulty caused by various factors, the results remained the same. Although DCE images were not acquired in about one-third of the MRIs, the diagnostic performance did not significantly differ whether DCE was added or not. This supports the view of the PI-RADS Steering Committee who set the role of DCE in the determination of PI-RADS V.2.1 Assessment Category secondary to T2WI and DWI because the adding value of DCE is not yet firmly established.[35] As for the specificity, a negative biopsy might not necessarily rule out the presence of small cancer lesions that are often detected in a surgical specimen. The diagnostic performance is based on the cases with pathological confirmation and is not free from verification bias. Whether to undergo a biopsy or not was eventually determined by the patients but not the MRI results, so the sensitivity could be falsely elevated or lowered if there were FN or TP cases among those who did not undergo a biopsy.

Generally speaking, a study on motivation is methodologically challenging because it has to be inferred from the behaviour or subjective reports in any investigation. The retrospective design also prevented us from performing an experiment to further test our hypothesis. Nevertheless, the specificity of our working condition—that is, the radiologists were under the same condition, but with a high degree of freedom for a part of their duties—offered a

golden opportunity to objectively quantify motivation as the volume of self-directed work. The retrospective approach enabled us to uncover the phenomena that were occurring in practice but that are less likely to be observable in an experimental condition, where participants would be more conscious of their behaviours.

CONCLUSION

In summary, our study showed that greater autonomy for diagnostic work, represented as the self-directed interpretive volume of CTs, was the factor that most contributed to improved prostate MRI interpretation performance. The lack of linear correlation between the performance and the assigned volume of prostate MRIs, per se, confirms the complex nature of human learning. Together, these findings support the hypothesis that successful promotion of internal drivers could have a pervasive positive impact on improving diagnostic performance.

ORCID iD
Shiori Amemiya http://orcid.org/0000-0002-2132-1642

REFERENCES

1. National Academies of Sciences, Engineering, and Medicine. *Improving diagnosis in health care*. Washington, DC: National Academies Press, 2015.
2. Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf* 2013;22:ii21–7.
3. Bruno MA, Walker EA, Abujudeh HH. Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. *RadioGraphics* 2015;35:1668–76.
4. Berlin L. Accuracy of diagnostic procedures: has it improved over the past five decades? *AJR Am J Roentgenol* 2007;188:1173–8.
5. Berlin L, Hendrix RW. Perceptual errors and negligence. *AJR Am J Roentgenol* 1998;170:863–7.
6. Kim YW, Mansfield LT. Fool me twice: delayed diagnoses in radiology with emphasis on perpetuated errors. *AJR Am J Roentgenol* 2014;202:465–70.
7. Lee CS, Nagy PG, Weaver SJ, *et al*. Cognitive and system factors contributing to diagnostic errors in radiology. *AJR Am J Roentgenol* 2013;201:611–7.
8. Kan L, Olivotto IA, Warren Burhenne LJ, *et al*. Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. *Radiology* 2000;215:563–7.
9. Beam CA, Conant EF, Sickles EA. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. *J Natl Cancer Inst* 2003;95:282–90.
10. Buist DSM, Anderson ML, Haneuse SJPA, *et al*. Influence of annual interpretive volume on screening mammography performance in the United States. *Radiology* 2011;259:72–84.
11. Buist DSM, Anderson ML, Smith RA, *et al*. Effect of Radiologists' Diagnostic Work-up Volume on Interpretive Performance. *Radiology* 2014;273:351–64.
12. Haneuse S, Buist DSM, Miglioretti DL, *et al*. Mammographic interpretive volume and diagnostic mammogram interpretation performance in community practice. *Radiology* 2012;262:69–79.
13. Kohn LT, Corrigan JM, Donaldson MS. *To err is human: building a safer health system*. Washington DC: National Academies Press (US), 2000.
14. Do H. *An organisation with a memory: report of an expert group on learning from adverse events in the NHS*. London: HM Stationery Office, 2000.
15. Larson DB, Donnelly LF, Podberesky DJ, *et al*. Peer feedback, learning, and improvement: answering the call of the Institute of medicine report on diagnostic error. *Radiology* 2017;283:231–41.
16. Weinreb JC, Barentsz JO, Choyke PL, *et al*. PI-RADS prostate imaging – reporting and data system: 2015, version 2. *Eur Urol* 2016;69:16–40.
17. Lauritzen PM, Andersen JG, Stokke MV, *et al*. Radiologist-initiated double reading of abdominal CT: retrospective analysis of the clinical importance of changes to radiology reports. *BMJ Qual Saf* 2016;25:595–603.
18. Abujudeh HH, Boland GW, Kaewlai R, *et al*. Abdominal and pelvic computed tomography (CT) interpretation: discrepancy rates among experienced radiologists. *Eur Radiol* 2010;20:1952–7.
19. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 2000;55:68–78.
20. Deci E, Vallerand R, Pelletier L, *et al*. Motivation and education: the Self-Determination perspective. *Educ Psychol* 1991;26:325–46.
21. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemp Educ Psychol* 2000;25:54–67.
22. Williams GC, Rodin GC, Ryan RM, *et al*. Autonomous regulation and long-term medication adherence in adult outpatients. *Health Psychology* 1998;17:269–76.
23. Williams GC, Freedman ZR, Deci EL. Supporting autonomy to motivate patients with diabetes for glucose control. *Diabetes Care* 1998;21:1644–51.
24. Ryan RM, Plant RW, O'Malley S. Initial motivations for alcohol treatment: relations with patient characteristics, treatment involvement, and dropout. *Addict Behav* 1995;20:279–97.
25. Deci EL. Effects of externally mediated rewards on intrinsic motivation. *J Pers Soc Psychol* 1971;18:105–15.
26. Glasziou PP, Buchan H, Del Mar C, *et al*. When financial incentives do more good than harm: a checklist. *BMJ* 2012;345:e5047.
27. McDonald R, Harrison S, Checkland K, *et al*. Impact of financial incentives on clinical autonomy and internal motivation in primary care: ethnographic study. *BMJ* 2007;334.
28. Woolhandler S, Ariely D, Himmelstein DU. Why pay for performance may be incompatible with quality improvement. *BMJ* 2012;345:e5015.
29. Roland M, Martin Roland D. Linking physicians' pay to the quality of care — a major experiment in the United Kingdom. *N Engl J Med* 2004;351:1448–54.
30. Marshall M, Harrison S. It's about more than money: financial incentives and internal motivation. *Qual Saf Health Care* 2005;14:4–5.
31. European Society of Radiology (ESR). Summary of the proceedings of the International Summit 2015: general and subspecialty radiology. *Insights Imaging* 2016;7:1–5.
32. Royal College of Radiologists. *Clinical radiology UK workforce census 2017 report*. London: Royal College of Radiologists, 2017.
33. Rosenkrantz AB, Wang W, Hughes DR, *et al*. Generalist versus subspecialist characteristics of the U.S. radiologist workforce. *Radiology* 2018;286:929–37.
34. Rosenkrantz AB, Ayoola A, Hoffman D, *et al*. The learning curve in prostate MRI interpretation: Self-Directed learning versus continual reader feedback. *AJR Am J Roentgenol* 2017;208:W92–100.
35. Turkbey B, Rosenkrantz AB, Haider MA, *et al*. Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. *Eur Urol* 2019;76:340–51.