

Received:
09 January 2019

Revised:
20 March 2019

Accepted:
21 March 2019

© 2019 The Authors. Published by the British Institute of Radiology under the terms of the Creative Commons Attribution 4.0 Unported License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Cite this article as:

Taylor-Phillips S, Stinton C. Fatigue in radiology: a fertile area for future research. *Br J Radiol* 2019; **92**: 20190043.

REVIEW ARTICLE

Fatigue in radiology: a fertile area for future research

SIAN TAYLOR-PHILLIPS, Ph.D and CHRIS STINTON

Warwick Medical School, The University of Warwick, Coventry, England

Address correspondence to: Dr Sian Taylor-Phillips
E-mail: s.taylor-phillips@warwick.ac.uk

ABSTRACT

Fatigue in radiologists may be responsible for a large number of medical errors. This review describes the latest research on fatigue in radiology. This includes measurement methods, and recent evidence on how fatigue affects accuracy in laboratory test conditions and in clinical practice. The extensive opportunities for future research in the area are explored, including testing interventions to reduce fatigue-related error, and further understanding of which fatigue measures correlate with errors. Finally we explore the possibility of answering these questions using large population-based observational studies and pragmatic integrated randomised controlled trials.

INTRODUCTION

Fatigue in medical professionals is detrimental to wellbeing and morale, and may result in medical errors. In this review, we focus on the effect of fatigue on medical errors by radiologists. There may be more than 11,000 preventable hospital deaths in English NHS acute hospitals each year of which diagnostic error contributes to over eight thousand.¹ This is a conservative estimate, with previous estimates serious disability or death caused by healthcare interventions in as many as 60 000 to 255 000 NHS patients.² The proportion of these which can be attributed to fatigue is not known. However we do know that more radiological errors occur towards the end of a shift than at the beginning, and that longer shifts in medicine are associated with higher rates of medical errors, and personal injury.³⁻⁵ The problems are exacerbated because the workload of radiological services is increasing dramatically (by an estimated 10–12% per year) and this is occurring without a concomitant increase in the number of staff.⁶ While the potential harm from fatigue in radiology is very large, there is surprisingly little research in the area. In this review we examine the different types and classifications of fatigue, the current evidence base using laboratory test sets and in real world observational studies and trials, and discuss future research directions.

MEASUREMENT OF FATIGUE

Fatigue is commonly described as feelings of weakness, lack of energy, and a desire to rest, and is associated with impairments in the ability to function.^{7,8} It is typically divided into two related subtypes: physical and mental fatigue. Physical fatigue is a deterioration in the muscles' ability to create

or sustain force, which results in difficulties in controlling and co-ordinating muscles, while mental fatigue is a reduction in the ability to perform mental tasks.⁹ In relation to radiology there is further subset: ocular fatigue related to vision. Measurement of fatigue can be objective or subjective, and each measurement method may cover one or more of these categories.

A range of tools have been developed to measure fatigue. It is of note that some measures used to examine fatigue were designed to measure tiredness. It much of the literature, fatigue and tiredness (or 'sleepiness') and used interchangeably. Several reviews have examined these tools.^{8,10,11} The majority of these rely of subjective assessment in the form of self-reported rating scales or checklists. Shahid and colleagues described both short-term (e.g. Karolinska Sleepiness Scale,¹² Stanford Sleepiness Scale¹³ and global measures of sleepiness (e.g. Epworth Sleepiness Scale,¹⁴ Sleep Wake Activity Inventory.¹⁵ Krupinski and Reiner reported on several additional measures, including the Pittsburgh Sleep Quality Index¹⁶ and the Swedish Occupational Fatigue Inventory (SOFI).^{10,17}

Objective measures of fatigue have also been described. For example, the Maintenance of Wakefulness Test (which examines alertness during the day) in which participants sit in a dimly lit room and are instructed to remain awake; the Psychomotor Vigilance Task (a test of sustained attention, during which participants are instructed to respond to stimuli by press a button – the test measures the number of responses/failures to respond); and the Continuous Performance Test (another measure of sustained attention, during

which participants are presented with unstimulating tasks and are instructed to respond to certain stimuli while ignoring others). Krupinski, Reiner, Waite and colleagues have described alternative approaches to the measurement of fatigue, such as monitoring of blood pressure, galvanic skin responses and heart rate.^{10,11} They also highlight the importance of visual fatigue in medical imaging. Interpreting medical images is a demanding and repetitive visual search task, and errors (false-negatives) are relatively common.¹⁸ Aside from self-reported measures, the main approaches to examining visual fatigue are through the assessment of accommodation (the adjustment of the curvature of the lens of the eye to change focus between objects in the distance and up close) and vergence (obtaining/maintaining binocular vision through the movement of eyes in opposite directions). Accommodation and vergence decline with fatigue.^{10,11} Others suggest critical fusion frequency (the frequency at which a flickering light is viewed as stable) and eye blink rate as objective measures of visual fatigue.^{19,20} Few of these measures have been employed in research on fatigue in medical imaging.

While there is reasonable consensus among researchers about the range of objective and subjective tools which are suitable to measure fatigue, there is very little research which links scores on these measures to clinically relevant outcomes from radiology, such as detection rates or error rates. Without knowing the relationship of these fatigue measures to clinical outcomes, studies using these measures are less likely to influence clinical practice. While it is important to minimise fatigue for the wellbeing of radiologists, understanding how fatigue can result in missing clinically relevant abnormalities may be more influential. Therefore, patterns of test accuracy, either measured continuously over the course of a radiology reading session or over the whole radiology work day, are often used as an alternative objective measure of fatigue. The psychology literature contains an abundance of studies examining the accuracy of novices at search and vigilance tasks typically lasting up to an hour.²¹ This field of research stemmed from an observation that sensitivity of radar operators to detect enemy aircraft or ships on radar screens decreased with time on task, called the vigilance decrement.²² The radiology reading task is considered a similar high intensity low signal salience task which may exhibit similar patterns,²³ (although differences in expertise, task, and task importance may limit generalisability). Also well-established from the field of psychology is the prevalence effect, where radiologists sensitivity to detect abnormalities is lower when examining test sets with lower prevalence of abnormalities.²⁴ This may be driven by radiologists expectations of the probability that a case is abnormal before looking at it, resulting in decreases in sensitivity with time on task in low prevalence radiology reading.²⁵ There is also emerging research on visual adaptation to examining radiology images, with further exploration required to determine whether this impacts patterns of accuracy with time on task.²⁶ There is a body of research about circadian rhythms (that being, biological processes that vary rhythmically over the course of, approximately, 24 h). In laboratory settings, time-related variations in performance have been observed for a range motor-, visual-, verbal-, memory-, and sensory tasks.²⁷ In general, performance on tasks improves over the course of the day and declines during

the night, albeit with fluctuations.²⁷ There is some evidence of a dip in performance after lunch.²⁸ Circadian rhythms have been hypothesised to apply to radiology accuracy, with some studies reporting that diagnostic accuracy is worse when the interpretation of images takes place later in the day than when it takes place earlier in the day.^{29,30} Other studies do not support this.^{31,32}

RECENT RESEARCH IN THE LABORATORY SETTING

Studies of the impact of fatigue on medical imaging are rare. Stec and colleagues conducted a systematic review looking for published research on the topic up to January 2017.³³ They identified 27 relevant papers, fewer than half of which were primary studies ($n = 10$) or 'other' types of studies ($n = 3$). The remaining 14 papers were themselves reviews.

Krupinski et al conducted a series of experimental studies.^{29,30,34,35} In the first of these studies, detection of bone fractures was assessed in 40 readers (20 radiologists and 20 residents) who read a series of 60 cases, comprising 2–4 images, before (early condition) and after (late condition) a day of routine clinical practice.²⁹ General fatigue was assessed using the SOFI, and visual fatigue (eye strain) was assessed using WAM-5500 Auto Refkeratometer (Grand Seiko, Hiroshima, Japan), which collects pupil diameter and refractive measurement. The authors found that subjective reports of fatigue and eye strain were significantly greater at the 'late' time point compared to the 'early' time point, and that diagnostic accuracy was significantly lower after a day of clinical practice than before (area under the curve early = 0.885, late = 0.852). No direct assessment was made between fatigue measurements and task performance.

In the second of their studies, 44 readers (22 attending radiologists and 22 radiology residents) sought to identify pulmonary nodules in CT of the chest.³⁰ As in their previous study, performance was examined before and after a day of clinical practice. General fatigue was assessed using the SOFI, with dark vergence (convergence of the eyes in the absence of stimuli) used as a measure of eye strain. Similar results were observed to those of their earlier study²⁹; diagnostic accuracy was significantly worse at the end of the day compared to the beginning of the day for the radiology residents (79% vs 75%). No such difference was observed for the attending radiologists. General and visual fatigue were not consistently higher in the late group. Degree of fatigue (*i.e.* score on the fatigue inventory or degree of visual strain) and task performance was not directly assessed.

In the third study, 20 radiologists examined computer tomography of multiple injuries after they had completed 8 h of clinical work (the so-called 'fatigued' group). Fatigue was measured as per the study of Krupinski et al.²⁹ The fatigued participants were matched to 20 non-fatigued participants who had taken part in previous studies by the same research group.^{29,30} For major fractures, there was no difference in detection between the fatigued and non-fatigued group (ROC AUC 0.945 vs 0.944, respectively). Again, no direct assessment was conducted between levels of fatigue and task performance. These studies provide some indication that radiologist expertise and experience may affect

amount of fatigue experienced, and/or moderate the relationship between fatigue and accuracy, which merits further investigation.

Finally, Hanna et al examined the effect of overnight shifts on diagnostic accuracy and fatigue in 12 radiologists (five faculty members and seven radiology residents).³⁵ The participants viewed bone radiographs during two conditions, once during the day (non-fatigued condition) and once in the morning after an overnight shift (fatigued condition). They also completed the SOFI. There were significant differences between the two conditions: there were higher fatigue scores on each of the domains of the SOFI, and lower diagnostic accuracy (0.806 vs 0.926) after the shift.

The results of these studies is supported by the review of Stec et al, which concluded that visual fatigue is relative common in radiologists (approximately 35% reported eye strain), and that visual and physical fatigue are associated with worse diagnostic accuracy (as measured by performance during extended periods of time).³³

RECENT RESEARCH IN CLINICAL PRACTICE

While the majority of evidence on the effect of fatigue on performance in medical imaging has been derived from experimental studies, a small number of studies have taken place in clinical practice. This is particularly valuable for understanding the importance of fatigue in the 'real world' as prior evidence has suggested variable associations between performance on experimental tasks and that observed in imaging in clinical practice.^{36,37}

Ruutiainen and colleagues compared discrepancies between the preliminary and final interpretation of 8062 medical images of 10 radiology residents who were working long hours (more than 10 consecutive hours). They found that there was a significantly higher rate of major discrepancies (which potentially had an impact on patients) in reports from the final 2 h of the residents' shifts compared to earlier shifts (2% compared to 1%).

Hanna et al retrospectively analysed the effects of shift length, schedule (regularly, holiday, weekend, extra, backup), and volume of workload on the accuracy of interpretation in a large sample ($n = 2,922,377$) of radiologic examinations covering a broad range of medical specialties.⁵ They found that both longer shifts and higher volumes of work (which are linked to fatigue) were associated with a greater number of major discrepancies in the interpretation of images. Further, a greater number of major discrepancies occurred during the latter part of shifts.

To the best of our knowledge, there has only been one prospective study of fatigue in medical imaging, the Changing case Order to Optimise patterns of Performance in Screening (CO-OPS) trial. This was a large randomised controlled trial that included data from over one million women. Taylor-Phillips and colleagues examined the impact of the vigilance decrement on the performance of pairs of readers (radiologists, radiography advanced practitioners, and breast clinicians) who were evaluating mammograms in the English breast cancer screening programme.^{38,39} The reader pairs interpreted batches

of approximately 35 mammograms in succession. For half of the pairs, the readers reviewed the batches in the same order (the control group), and for the other half of the pairs one reader read the batches forward, and the other read the batches backwards (the intervention group). If the vigilance decrement was in evidence, the cancer detection rate should decrease with time on task in the control group (as by the end of the batch both readers would be in a low vigilance state) but not in the intervention group (as the readers will be in low vigilance states at different points in the batch). However, no difference was observed in cancer detection rates between the two groups (OR, 1.01; 95% CI, 0.97–1.06), indicating no vigilance decrement. Interestingly, recall rate decreased over time, suggesting that reader performance in terms of positive predictive value actually increased with time on task.

Further analysis of a subset of the CO-OPS dataset has been reported by Stinton et al.³⁸ They examined recall and cancer detection rates by readers over the course of the day, with data were divided into three equal time periods, based on when the interpretations of mammograms were carried out: 9am to 4:59pm, 5pm to 12:59pm, and 1am to 8:59am. Variability in both recall and cancer detection rates were observed over the course of the day, and multilevel logistic regression indicating that females whose mammograms were interpreted between 5pm and 12:59pm were 1.07 (95% credible interval 1.03–1.11) times more likely to be recalled than females whose mammograms were read during each of the other time periods ($p < 0.001$). There was no significant association between time period and cancer detection rate. These patterns were observed in analyses with (1) all readers included, and (2) excluding readers who didn't work during the evening.

FUTURE RESEARCH OPPORTUNITIES

There is good evidence that fatigue is a serious issue affecting error rates across medicine including radiology. Future research is required to understand which interventions can reduce fatigue in radiologists, and in turn reduce the consequent medical errors. Further, there are many subjective and objective measures of fatigue, but research is required to identify which ones are associated with fatigue-related medical errors. Such information could be used to select fatigue measurements which would identify states of fatigue that may lead to medical errors in real-time. Elsewhere in radiology, technological advances associated with artificial intelligence may affect several elements of the radiology task including those driving fatigue, and such associations require investigation. Finally, the advent of big data provides an opportunity to answer these research questions using methods that were not previously widely available, in large-scale observational and integrated pragmatic trial studies.

There is a scarcity of research investigating the effectiveness of interventions to reduce fatigue and fatigue-related errors in radiology. The largest study was a randomised controlled trial of an intervention to change case order in mammography screening detailed above.³⁹ This study found that the intervention was not effective. The effect of other potential interventions to reduce fatigue in a single reading session, such as break scheduling,

interruptions, caffeine, workstation design and ambient lighting, are yet to be robustly tested in a real-world setting. Perhaps the greater opportunity is in investigating fatigue towards the end of a long shift, which may be greater than that which develops within a single radiology session. Interventions in this context may be more organisational such as scheduling of tasks within a shift so that safety critical tasks are towards the beginning, or changes to shift length, shift times, or staffing levels. All such interventions should be designed to support radiologists in managing their own fatigue, rather than taking a top-down mandated approach, and so necessarily study designs to investigate these must be pragmatic and include adaptation to the local context.

We do not know which fatigue measurements are related to increases in rates of medical errors. Large studies are required employing several simple fatigue measures in the same radiologists at the same time, and relating these to real-world error rates. These data would inform which fatigue measures can be used as a proxy for fatigued states that cause medical errors. Then the same measures could be used as proxy outcomes in research studies, and as useful feedback mechanisms in clinical practice.

The advent of artificial intelligence examining radiological images is not new, it is akin to computer aided detection, which has been implemented for many years with varying levels of success. What is new is developments in data storage, linkage and artificial intelligence are resulting in a new generation of automated image reading tools. The accuracy of these is not the focus of this review, but accuracy appears to be dependent on access to large numbers of images with known ground truth, which are becoming increasingly available. What is important from the perspective of radiologist fatigue, is how these developments interact with the radiologist. Fatigue research has shown that in some circumstances specificity and positive predictive value improve as the reading session progresses,⁴⁰ which may explain the mechanism through which batch reading improves specificity.⁴¹ Building on this human factors and fatigue research we can postulate that if artificial intelligence was used to sift out the straightforward normal cases then this would increase case difficulty read by radiologists, which may increase fatigue but also increase prevalence and therefore sensitivity.^{23,24} If artificial intelligence were implemented to highlight suspicious radiological areas then it may interrupt the reading session with false positive prompts, and negatively affect specificity, in a similar manner to its predecessor, computer-aided detection in breast screening.^{42,43}

The challenge in measuring radiologist fatigue is one shared by researchers measuring all elements of radiologists' performance and accuracy, that of statistical power. We want to know which equipment, workstations, working hours, conditions and tests enable radiologists to be as accurate as possible. We want the outcomes of our research studies to be clinically meaningful, for example measuring numbers of clinically significant missed cases/abnormalities. This requires either test sets enriched with unrealistic numbers and types of abnormalities, or very large studies. As there is reasonable evidence that radiologists' behaviour examining enriched test sets is not

generalisable to clinical practice,^{36,37} the future appears to lie in very large studies.

The era of 'big data' provides great potential and risk for research. Computerised patient records and images on a large scale provide fantastic research opportunities that are only beginning to be realised. Breast cancer screening in England is an example at the forefront of this. This screening programme has one single computer programme to manage patient records, so recording is relatively uniform across the country. This computer programme has regular data linkage to and from the national cancer registry which contains long term patient follow-up data. This has allowed a wealth of observational research, and more recently changes to the software to integrate randomisation. This internal randomisation, and follow up to clinically relevant outcomes allows us to automate the running of randomised controlled trials such as the CO-OPS trial of one million women³⁹ and the age extension trial of six million.⁴⁴ Such automation allows much larger trials at significantly lower cost. Where possible such pragmatic integrated randomised controlled trials provide the least biased form of fatigue research available. Where randomisation is not possible or practical, for example comparing different reading room designs, observational research remains valuable.

In the transition to use of big data routinely in radiology research and audit there is a real risk of inappropriate analyses leading to incorrect conclusions. There is now a plethora of observational data available in radiology, in the images themselves and the associated patient data. Observational study designs have many confounders and caution is required in drawing conclusions from such designs. The increased availability of large sample sizes could lead to increased publication bias through publication only when results are 'interesting'. For example in the UK an observational analysis of increased death rates of patients admitted to hospitals at weekend was used as rationale for seven day working.^{45,46} However, a driving factor of the effect is increased severity of illness in those admitted at the weekend, with the authors themselves stating that "to assume that they [extra deaths at weekend] are avoidable would be rash and misleading". This is one of the more prominent examples of the difficulties of ascribing cause and effect when only observational data are available. Research methods training should become more widely available to those who audit and analyse such data, to ensure analysis takes account of confounders and to prevent incorrect inference. In fatigue research, particular attention should be paid to individual differences between radiologists, for example through including radiologists in analyses as a random effect.

The success of future research in fatigue will depend heavily on the culture in radiology work environments. An individual radiologist's performance data have the potential to be a very powerful tool for constantly improving performance, if communicated accurately in a positive learning environment. Similarly, research can be completely anonymised with respect to both patients and radiologists, allowing large research studies with lower perceived risk. If such data are used in an

adversarial manner, as performance management, in a blame culture, or even in misconduct lawsuits then big data could be perceived as big brother. This would limit the willingness of radiologists to engage with big data in an open manner conducive to learning and improvement.

CONCLUSION

There is clear evidence that fatigue, particularly towards the end of a long shift, contributes to serious medical errors, and increases the risk of missing abnormalities on imaging. There is a research need to investigate which interventions could reduce fatigue-related errors, and which measures of fatigue are strong

early indicators of a fatigued state that may lead to medical errors. The advent of big data in healthcare provides an opportunity to undertake large studies to address these questions, but attention must be paid to addressing confounding.

ACKNOWLEDGEMENTS

Dr Taylor-Phillips and Dr Stinton are supported by the NIHR CLAHRC West Midlands. Dr Taylor-Phillips is supported by an NIHR Career Development Fellowship (CDF-2016-09-018). The opinions are those of the authors and not the NIHR, the NHS or the Department of Health and Social Care.

REFERENCES

- Hogan H, Healey F, Neale G, et al. Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. *J BMJ Quality & Safety* 2012;bmjqs-2012-001159.
- Chief Medical Officer. *An organisation with a memory: Report of an expert group on learning from adverse events in the NHS*. London: The Stationery Office; 2000.
- Lockley SW, Barger LK, Ayas NT, Rothschild JM, Czeisler CA, Landrigan CP, et al. Effects of health care provider work hours and sleep deprivation on safety and performance. *Jt Comm J Qual Patient Saf* 2007; 33(11 Suppl): 7–18. doi: [https://doi.org/10.1016/S1553-7250\(07\)33109-7](https://doi.org/10.1016/S1553-7250(07)33109-7)
- Barger LK, Cade BE, Ayas NT, Cronin JW, Rosner B, Speizer FE, et al. Extended work shifts and the risk of motor vehicle crashes among interns. *N Engl J Med* 2005; 352: 125–34. doi: <https://doi.org/10.1056/NEJMoa041401>
- Hanna TN, Lamoureux C, Krupinski EA, Weber S, Johnson J-O. Effect of shift, schedule, and volume on interpretive accuracy: a retrospective analysis of 2.9 million radiologic examinations. *Radiology* 2018; 287: 205–12. doi: <https://doi.org/10.1148/radiol.2017170555>
- RCo R. *Clinical Radiology UK workforce census 2014 report*. London: The Royal College of Radiologists; 2015.
- Pigeon WR, Sateia MJ, Ferguson RJ. Distinguishing between excessive daytime sleepiness and fatigue: toward improved detection and treatment. *J Psychosom Res* 2003; 54: 61–9.
- Shahid A, Shen J, Shapiro CM. Measurements of sleepiness and fatigue. *J Psychosom Res* 2010; 69: 81–9. doi: <https://doi.org/10.1016/j.jpsychores.2010.04.001>
- Xu R, Zhang C, He F, Zhao X, Qi H, Zhou P, et al. How physical activities affect mental fatigue based on EEG energy, connectivity, and complexity. *Front Neurol* 2018; 9: 915–15. doi: <https://doi.org/10.3389/fneur.2018.00915>
- Krupinski E, Reiner BI. Real-time occupational stress and fatigue measurement in medical imaging practice. *J Digit Imaging* 2012; 25: 319–24. doi: <https://doi.org/10.1007/s10278-011-9439-1>
- Waite S, Kolla S, Jeudy J, Legasto A, Macknik SL, Martinez-Conde S, et al. Tired in the Reading Room: The Influence of Fatigue in Radiology. *J Am Coll Radiol* 2017; 14: 191–7. doi: <https://doi.org/10.1016/j.jacr.2016.10.009>
- Akerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. *Int J Neurosci* 1990; 52(1-2): 29–37. doi: <https://doi.org/10.3109/00207459008994241>
- Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. *Psychophysiology* 1973; 10: 431–6. doi: <https://doi.org/10.1111/j.1469-8986.1973.tb00801.x>
- Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep* 1991; 14: 540–5. doi: <https://doi.org/10.1093/sleep/14.6.540>
- Rosenthal L, Roehrs TA, Roth T. The sleep-wake activity inventory: a self-report measure of daytime sleepiness. *Biol Psychiatry* 1993; 34: 810–20. doi: [https://doi.org/10.1016/0006-3223\(93\)90070-T](https://doi.org/10.1016/0006-3223(93)90070-T)
- Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989;(eds 28): 193–213. doi: [https://doi.org/10.1016/0165-1781\(89\)90047-4](https://doi.org/10.1016/0165-1781(89)90047-4)
- Ahsberg E. Dimensions of fatigue in different working populations. *Scand J Psychol* 2000; 41: 231–41. doi: <https://doi.org/10.1111/1467-9450.00192>
- Miglioretti DL, Smith-Bindman R, Abraham L, Brenner RJ, Carney PA, Bowles EJA, et al. Radiologist characteristics associated with interpretive performance of diagnostic mammography. *J Natl Cancer Inst* 2007; 99: 1854–63. doi: <https://doi.org/10.1093/jnci/djm238>
- Maeda E, Yoshikawa T, Hayashi N, Akai H, Hanaoka S, Sasaki H, et al. Radiology reading-caused fatigue and measurement of eye strain with critical flicker fusion frequency. *Jpn J Radiol* 2011; 29: 483–7. doi: <https://doi.org/10.1007/s11604-011-0585-7>
- Sheppard AL, Wolffsohn JS. Digital eye strain: prevalence, measurement and amelioration. *BMJ Open Ophthalmol* 2018; 3: e000146. doi: <https://doi.org/10.1136/bmjophth-2018-000146>
- Thomson DR, Besner D, Smilek D. A critical examination of the evidence for sensitivity loss in modern vigilance tasks. *Psychol Rev* 2016; 123: 70–83. doi: <https://doi.org/10.1037/rev0000021>
- Mackworth N. *Researches on the Measurement of Human Performance*. London, UK: HMSO; 1950.
- See JE, Howe SR, Warm JS, Dember WN, et al. Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin* 1995; 117: 230–49. doi: <https://doi.org/10.1037/0033-2909.117.2.230>
- Gur D, Rockette HE, Armfield DR, Blachar A, Bogan JK, Brancatelli G, et al. Prevalence effect in a laboratory environment. *Radiology* 2003; 228: 10–14. doi: <https://doi.org/10.1148/radiol.2281020709>
- Evans KK, Birdwell RL, Wolfe JM. If you don't find it often, you often don't find it: why some cancers are missed in breast cancer screening. *PLoS One* 2013; 8: e64366–e66. doi: <https://doi.org/10.1371/journal.pone.0064366>

26. Kompaniez-Dunigan E, Abbey CK, Boone JM, Webster MA. Visual adaptation and the amplitude spectra of radiological images. *Cogn Res Princ Implic* 2018; **3**: 3. doi: <https://doi.org/10.1186/s41235-018-0089-4>
27. Blatter K, Cajochen C. Circadian rhythms in cognitive performance: methodological constraints, protocols, theoretical underpinnings. *Physiol Behav* 2007; **90**(2-3): 196–208. doi: <https://doi.org/10.1016/j.physbeh.2006.09.009>
28. Monk TH. The post-lunch dip in performance. *Clin Sports Med* 2005; **24**: e15–23xi-xii. doi: <https://doi.org/10.1016/j.csm.2004.12.002>
29. Krupinski EA, Berbaum KS, Caldwell RT, Scharztz KM, Kim J. Long radiology workdays reduce detection and accommodation accuracy. *J Am Coll Radiol* 2010; **7**: 698–704. doi: <https://doi.org/10.1016/j.jacr.2010.03.004>
30. Krupinski EA, Berbaum KS, Caldwell RT, Scharztz KM, Madsen MT, Kramer DJ, et al. Do long radiology workdays affect nodule detection in dynamic CT interpretation? *J Am Coll Radiol* 2012; **9**: 191–8. doi: <https://doi.org/10.1016/j.jacr.2011.11.013>
31. Al-s'adi M, McEntee M, Ryan E. Time of day does not affect radiologists' accuracy in breast lesion detection. *Proc SPIE Med Imag* 2011; **7966**: 1–7.
32. Cowley HC, Gale AG. Time of day effects on mammographic film reading performance. 1997;.
33. Stec N, Arje D, Moody AR, Krupinski EA, Tyrrell PN. A systematic review of fatigue in radiology: is it a problem? *AJR Am J Roentgenol* 2018; **210**: 799–806. doi: <https://doi.org/10.2214/AJR.17.18613>
34. Krupinski EA, Scharztz KM, Van Tassel MS, Madsen MT, et al. Effect of fatigue on reading computed tomography examination of The multiply injured patient. *J Med Imaging* 2017; **4**: 1. doi: <https://doi.org/10.1117/1.JMI.4.3.035504>
35. Hanna TN, Zygmunt ME, Peterson R, Theriot D, Shekhani H, Johnson J-O, et al. The effects of fatigue from overnight shifts on radiology search patterns and diagnostic performance. *J Am Coll Radiol* 2018; **15**: 1709–16. doi: <https://doi.org/10.1016/j.jacr.2017.12.019>
36. Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, et al. The "laboratory" effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. *Radiology* 2008; **249**: 47–53. doi: <https://doi.org/10.1148/radiol.2491072025>
37. Miglioretti DL, Ichikawa L, Smith RA, Buist DSM, Carney PA, Geller B, et al. Correlation between screening mammography interpretive performance on a test set and performance in clinical practice. *Acad Radiol* 2017; **24**: 1256–64. doi: <https://doi.org/10.1016/j.acra.2017.03.016>
38. Stinton C, Jenkinson D, Adekanmbi V, et al. Does time of day influence cancer detection and recall rates in mammography? *Proc SPIE 10136, Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment* 101360B 2017 10.1117/12.2254280.
39. Taylor-Phillips S, Wallis MG, Jenkinson D, Adekanmbi V, Parsons H, Dunn J, et al. Effect of using the same vs different order for second readings of screening mammograms on rates of breast cancer detection: a randomized clinical trial. *JAMA* 2016; **315**: 1956–65. doi: <https://doi.org/10.1001/jama.2016.5257>
40. Wolfe JM, Van Wert MJ. Varying target prevalence reveals two dissociable decision criteria in visual search. *Curr Biol* 2010; **20**: 121–4. doi: <https://doi.org/10.1016/j.cub.2009.11.066>
41. Burnside ES, Park JM, Fine JP, Sisney GA. The use of batch reading to improve the performance of screening mammography. *AJR Am J Roentgenol* 2005; **185**: 790–6. doi: <https://doi.org/10.2214/ajr.185.3.01850790>
42. Fenton JJ, Abraham L, Taplin SH, Geller BM, Carney PA, D'Orsi C, et al. Effectiveness of computer-aided detection in community mammography practice. *J Natl Cancer Inst* 2011; **103**: 1152–61. doi: <https://doi.org/10.1093/jnci/djr206>
43. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med Overseas Ed* 2007; **356**: 1399–409. doi: <https://doi.org/10.1056/NEJMoa066099>
44. ISRCTN Register Evaluating the age extension of the NHS breast screening Programme—trial registration. 2010;.
45. Freemantle N, Ray D, McNulty D, Rosser D, Bennett S, Keogh BE, et al. Increased mortality associated with weekend hospital admission: a case for expanded seven day services? *BMJ* 2015; **351**: h4596. doi: <https://doi.org/10.1136/bmj.h4596>
46. Rimmer A. The BMJ paper and seven day services.. In: ed) *BMJ (Clinical research)*. **352**; 2016. pp. i1193.