Sleep Medicine Reviews 46 (2019) 97–107

Contents lists available at ScienceDirect

Sleep Medicine Reviews

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

CLINICAL REVIEW

A review of current approaches for evaluating impaired performance in around-the-clock medical professionals

Coen R.H. Huizinga [a, 1], Fokkedien H.M.P. Tummers [a, 1], Perla J. Marang-van de Mheen [b], Adam F. Cohen [a], Koen E.A. van der Bogt [c, d, *]

[a] Centre for Human Drug Research, Leiden, the Netherlands
[b] Leiden University Medical Centre, Medical Decision Making, Leiden, the Netherlands
[c] Department of Surgery, Haaglanden Medical Centre, The Hague, the Netherlands
[d] Department of Surgery, Leiden University Medical Centre, Leiden, the Netherlands

ARTICLE INFO

Article history:
Received 1 May 2018
Received in revised form
19 March 2019
Accepted 10 April 2019
Available online 25 April 2019

Keywords:
Fatigue
Sleep deprivation
Performance
Healthcare professionals
Review
Evaluation
Assessment

SUMMARY

The need for data to study the relationship between fatigued healthcare professionals and performance outcomes is evident, however, it is unclear which methodology is most appropriate to provide these insights. To address this issue, we performed a systematic review of relevant articles by searching the MEDLINE, EMBASE, Cochrane, Web of Science, and CINAHL databases. The literature search identified 2960 unique references, of which 82 were identified eligible. The impact on performance was studied on clinical outcomes, medical simulation, neurocognitive performance, sleep quantification and subjective assessment. In general results on performance are conflicting; impairment, no effect, and improvement were found. This review outlines the various methods currently available for assessing fatigue-impaired performance. The contrasting outcomes can be attributed to three main factors: differences in the operationalisation of fatigue, incomplete control data, and the wide variety in the methods used. We recommend the implementation of a clinically applicable tool that can provide uniform data. Until these data become available, caution should be used when developing regulations that can have implications for physicians, education, manpower planning, and – ultimately – patient care.

Introduction

The effect of extended working hours and fatigue on a physician's performance has long been a topic of interest. Although both mental and physical exertion are believed to adversely affect cognitive functioning, there is currently no consensus regarding the impact on patient safety [1]. A tired physician can be prone to commit negligence and/or errors that can lead to complications [2], thereby potentially compromising patient safety [3]. For example, surgeons have reported that lapses in judgement are the largest contributing factor to surgical error, and 8.9% of surgeons reported that they made a major medical error within the previous three months [4].

In addition, a physician's well-being can be negatively affected by long working hours and/or sleep deprivation. Studies indicate that sleep-deprived or distressed physicians have an increased risk of motor vehicle accidents, as well as an increase in somatic complaints, depression, and burnout [5–7]. This finding is particularly alarming for around-the-clock healthcare professionals, who generally work long hours under constant high pressure to make correct decisions.

Concerns regarding patient safety and the clinical resident's well-being led to the implementation of restrictions for residents' working hours by the US Accreditation Council for Graduate Medical Education (ACGME) and the European Working Time Directive, which limit a resident's number of continuous on-duty hours to 80 [8] and 48 [9] hours, respectively. This large difference between US and European guidelines in itself already provides a sound argument that a consensus is urgently needed with respect to interpreting current data regarding the effect of long working hours. For example, sleepiness cannot be defined simply in terms of the total amount of sleep lost [10], and limiting a physician's working hours does not necessarily result in a physician who is well-rested [11]. Furthermore, such limitations do not apply to attending surgeons, as they are still permitted to perform elective surgeries – possibly in a sleep-deprived state – following an overnight call session [12,13].

* Corresponding author. Dept. of Surgery, Leiden University Medical Centre, PO Box 9600, 2300, RC, Leiden, the Netherlands.
E-mail address: K.E.A.van_der_Bogt@lumc.nl (K.E.A. van der Bogt).
[1] Both Huizinga and Tummers contributed equally.

https://doi.org/10.1016/j.smrv.2019.04.007
1087-0792/

Many professionals who work around the clock under highly demanding working conditions (for example aviators) have a Fatigue Risk Management system available [14]. In contrast, however, no "fit-to-perform" standards are currently available for medical professionals, making an objective assessment of their impaired performance a precarious task. Nevertheless, the need for such standards in medicine is clear, particularly given that physicians are generally unable to accurately self-assess their fitness to perform [15,16]. Moreover, disagreement has been reported between subjective (i.e., self-reported) and objective assessments of sleepiness, particularly while working an overnight on-call shift, and this is believed to pose a significant risk to the safety of healthcare professionals and their patients [17].

Because no standards are currently available for healthcare professionals, studies lack a comprehensive range of performance parameters; consequently, a relevant frame of reference for determining what is considered acceptable is lacking. In addition, because established cut-off values are non-existent, terms such as "fatigued" and "fit" in the context of the healthcare professional remain to be defined. Taken together, the need for data regarding the impaired performance of a mentally or physically exerted physician is evident; however, it is currently unclear which approach is most appropriate for providing these data.

To address these issues three primary objectives were established. Our first objective in conducting this review was to prepare an inventory of all current tools (e.g., methods, tests, and parameters) that have been used to assess the effects of decreased alertness and sleep deprivation on the performance of healthcare professionals. Our second objective was to categorise all neurocognitive tools that are used in specific domains and are believed to be important for evaluating a physician's performance. Our third and final objective was to offer a recommendation for an assessment approach that is appropriate for use in a clinical setting and may lead to a more personalised approach to managing fatigue and identifying risk factors. Additionally, we provide an overview of the wide range of study results of the included studies that assessed effects of fatigue and sleep deprivation on performance of healthcare professional.

Methods

Search strategy and data sources

A systematic search was conducted in the MEDLINE, EMBASE, Cochrane, Web of Science, and CINAHL databases (see supplemental material). There was no restriction with respect to date of publication; however, the search was limited to publications written in English. The reference lists of the included articles were also screened in order to identify any additional publications that were not identified in the original search. First, two authors (CH and KvdB) independently assessed the title and abstract of each article identified by the search, and the full text of each relevant article was obtained. Next, one author (CH) excluded articles based on the full text, after which a second author (KvdB) verified that the selected articles met the inclusion criteria. Where necessary, a third or fourth author (FT and AC) was available for consultation in order to reach a consensus.

Inclusion process

The following inclusion criteria were used: the study had to be published in a peer-reviewed journal; the study population consisted of at least ten physicians and/or residents; and the outcomes of interest were objective and/or subjective methods used to investigate the effects of fatigue and/or sleep deprivation on the subjects' performance. Studies that solely used non-validated questionnaires to assess fatigue were excluded. In addition, opinion papers, editorials, and articles deemed to be irrelevant (e.g., papers that focussed solely on work-hour regulations or on stress and/or burnout) were excluded based on an assessment by two authors (CH and KvdB).

Scoring of measurement methods

Diagnostic methodology was assessed using the criteria for developing diagnostic tests formulated previously by West et al. [18] The following five essential test domains were evaluated: study population, test description, appropriate reference standard, blinded comparison, and avoidance of verification bias. The majority of diagnostic studies did not comply with all five criteria; therefore, a tripartite evaluation ("yes", "partial", or "no") was used (supplemental table 1). Studies were included in the review when at least two criteria were met. For each included study, the following data were extracted: type of clinical event or assessment method, study contrast, study population, study design, test system, blinding, quantification of mental of physical exertion, number of cases, and outcomes.

Neurocognitive tools

All neurocognitive tools in the articles were categorised into one or more of the following six key neurocognitive domains [19]: visuospatial perception and constructional skills; executive functioning (e.g., decision making, planning, and reasoning); attention and concentration; mood, personality, and behaviour; memory (including working memory); and speech, language, and communication skills.

Definition of fatigue

In order to create an inventory of all explorations in which the effects of fatigue are assessed, a standardised definition of the word "fatigue" was required, as fatigue was defined differently among the included studies. To provide an overview, we established the following categories for describing fatigue: "daytime vs. night time", "post-call", "fatigued", and "sleep-deprived". The category "daytime vs. night time" describes the contrast between results obtained during the day and results obtained during the night. The category "post-call" describes results that compared outcomes obtained post-call versus pre-call, on-call, or at the end of the day shift. The category "fatigued" describes the contrast between results obtained at the beginning of the day versus the end of the day (i.e., duration of shift) attributed to mental and/or physical exertion. Lastly, the category "sleep-deprived" describes studies that measure fatigue by the quantity of recent sleep or studies that define cohorts a sleep-deprived.

Results

Included studies

Our initial literature search identified 2960 unique articles, 82 of which were ultimately determined to be eligible (Fig. 1). These studies assessed performance using both objective and subjective outcome measures. Twenty-nine studies assessed clinical outcomes such as surgical or clinical errors, diagnostic errors, medication or administrative errors, and adverse events experienced before, during, or after hospitalisation (Supplemental table 2). Forty-five studies assessed performance individually and objectively using medical simulations (n = 16 studies, Supplementable

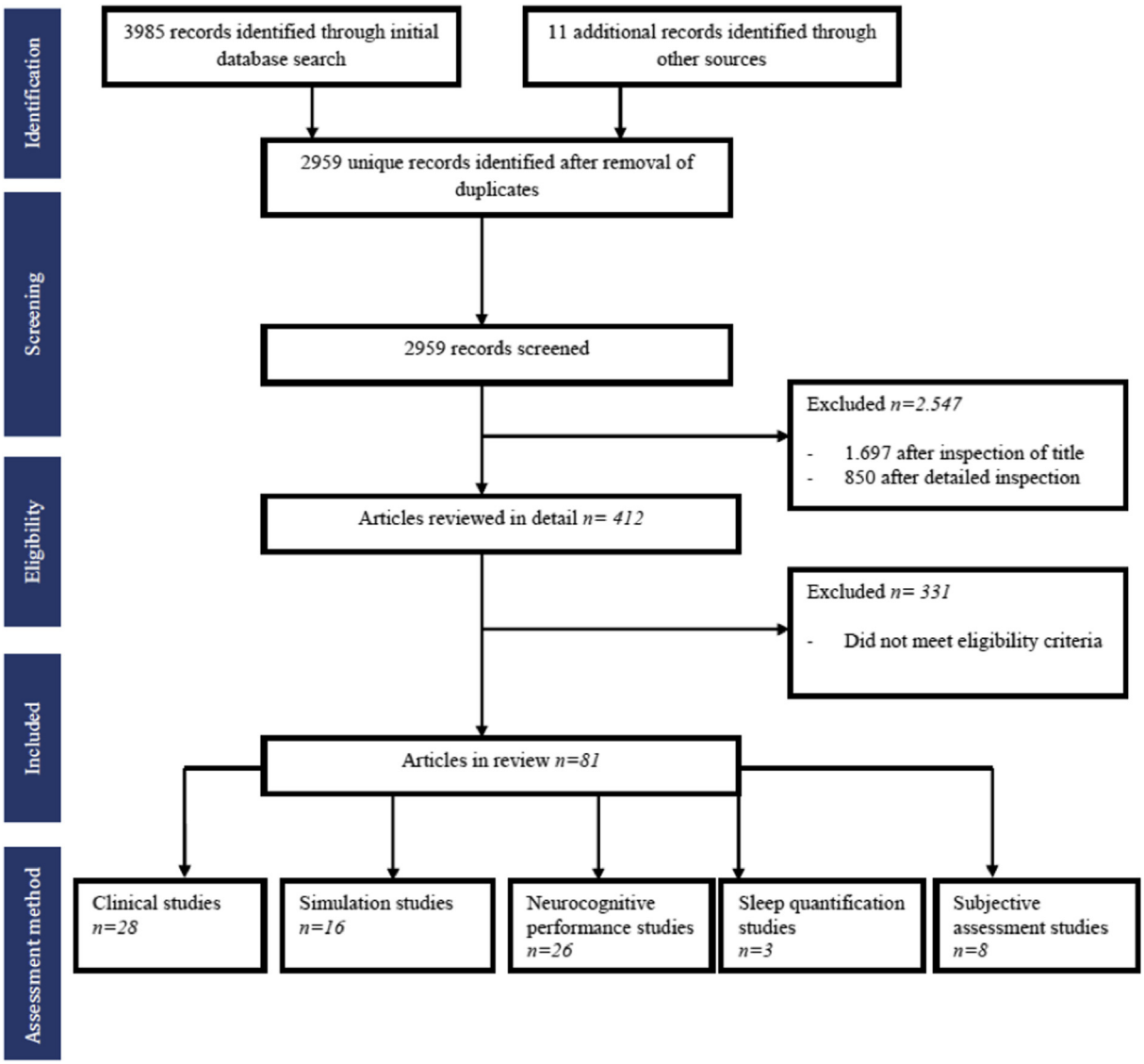

Fig. 1. Flow chart depicting the search and inclusion/exclusion criteria.

table 3), neuropsychological and motor function tests (n = 26 studies, Supplemental table 4), or by quantifying sleep (n = 3 studies, Supplemental table 5). Finally, eight studies assessed performance subjectively using a questionnaire (n = 8 studies, Supplemental table 6).

Clinical studies

In total, 29 studies (17 surgical and 11 non-surgical) assessed the association between fatigued physicians and clinical outcomes (Supplemental table 2).

Surgical outcomes

Three studies reported the correlation between surgical complications and mental and/or physical exertion with regard to the length of shift cycle. Two of these studies found an association between a later surgery start time and the number of surgical errors [2,20], whereas the third study found no such association [15]. Six observational cohort studies compared night time surgery with daytime surgery; three studies reported an increased incidence of complications associated with night time surgery [4,21,22], whereas the other three studies found no negative effects of night time surgery [17,23–25]. Eight control-matched studies were also included. Six of these studies matched procedures performed by post-call surgeons with procedures performed by control surgeons who had not been on-call [17,26–30], and two studies matched procedures performed by the same surgeon under different conditions [31,32]; no difference in the number of complications was observed. Lastly, one study examined the relationship between sleep deprivation and complications associated with placement of an epidural during labour and found no correlation [33].

The current data is conflicting, among others due to the diversity in methodology and great variety of covariates, and does not seem to be sufficient to provide clear insights regarding the relationship between physician impairment and surgical outcomes. Surgical outcomes in general (especially clinically relevant complications) are a result of team effort and patient characteristics and can only be regarded as a reliable outcome if recorded prospectively in detail and by multiple observers, as was the case in one study [22]. However, this study used time of day as the influencing factor and as such does not provide information about individual sleepiness of the surgeon making the error .

Non-surgical outcomes: diagnostic, administrative, and medication errors

Eleven non-surgical studies were included in our analysis (Supplemental table 2). One study compared night time and daytime cases and found no difference with respect to either the number of errors committed or accuracy in performance [34]. Four studies examined fatigued-related errors (e.g., mental and/or physical exertion); three of these studies found an association between length of shift cycle and the number of diagnostic errors [35–37], and one study found no association between length of shift cycle and medication errors [38]. Three studies assessed the effects of fatigue on diagnostic errors, with conflicting results. Krupinski et al. reported that significantly more errors occurred among radiologists after a day of reading exams [12], whereas another publication by the same group reported impaired performance among residents but not among specialists [39]. In contrast, Christensen et al. reported no significantly impaired performance among fatigued residents after working at least fifteen consecutive hours [40]. Two studies compared post-call physicians with rested controls. Hendey et al. found that more medication errors occurred during post-call shifts [14], whereas Amirian et al. found no significant difference with respect to the quality of medical records [41]. Lastly, Landrigan et al. conducted a prospective randomised controlled study and found that interns had a 5.6-fold higher rate of serious medical errors when working frequent $\geq$24-hour shifts compared to when they worked shorter shifts; this increase in errors included a 21% increase in the number of serious medication errors [1].

Whether an association was found depended on which groups were compared and how fatigue was operationalised; studies that assessed length of shift cycle or sleepiness were more likely to find a negative effect compared to studies that examined different types of work shifts. Moreover, the effect of confounding factors with respect to clinical outcome must also be considered; factors such as patient characteristics and the physician's experience were included in seven surgical studies [4,15,24,25,28,31,32] and nine non-surgical studies [1,12,34,35,37–41], respectively. The effect of such covariates has yet to be investigated but this is unlikely to provide valuable results when current data is meta-analyzed [42].

Non-clinical studies

In addition to clinical outcomes, the relationship between physical or mental exertion and impaired performance has also been examined in a broad range of non-clinical studies. These studies include simulations, measures of neurocognitive performance, and the quantification of sleep data in order to assess the impact on performance.

Simulation studies

Laparoscopic surgery is physically demanding for the surgical staff, and performance is believed to decline significantly when the surgeon is fatigued. Approximately 20 years ago, Taffinder et al. reported that performing laparoscopic surgery post-call resulted in a longer procedure and more errors [43]. Since then, a number of validated surgical simulators have been used to measure the effects of physical or mental exertion and sleep deprivation on both technical and non-technical skills. Sixteen studies involving simulators were included in our analysis (Supplemental table 3); thirteen of these studies reported the outcome of a laparoscopic simulation, and three studies reported non-technical or non-surgical skills.

Eleven of the thirteen laparoscopic studies assessed post-call performance. Five of these eleven studies reported a significant decline in post-call performance [44–48], four reported no change in post-call performance [49–52], and two reported improved post-call performance [53,54].

The two remaining laparoscopic studies measured fatigued and sleep-deprived surgeons. One study reported impaired performance [55], whereas the other study reported no change in performance [56].

The three other simulation studies assessed either non-technical or non-surgical skills. Gordon et al. reported significant improvements in post-call performance on an acute care simulator when the overnight shift was shortened from 24 to 30 hours to a 16-hour shift [57]. Murray et al. reported significantly impaired post-call performance on a driving simulator performance due to sleep disruptions [58]. Finally, Storer et al. studied the effects on various shift lengths on post-call skills while performing various clinical procedures, including endotracheal intubation, venous cannulation, and umbilical artery catheterisation, but found inconsistent results [59]. Whether an association was found – and the nature of that association – appeared to depend on the study's operationalisation of fatigue.

Neurocognitive performance studies

In contrast with simulator studies, performance has been evaluated using a wide range of neurocognitive and psychomotor tests. More than 30 years ago, Leighton et al. reported a significant decline in test performance after an on-call night [60]. Since then, researchers have used methods to assess neurocognitive functioning, focussing on attention, working memory, dexterity, and fine motor skill to evaluate changes in performance. In total, 26 studies were eligible for inclusion in our review.

Twelve of these studies confirmed the initial results reported by Leighton et al. with respect to significant impairment of key neurocognitive domains in post-call physicians [61–69], fatigued physicians [70] and sleep-deprived physicians [71,72]. In contrast, seven studies did not observe any significant changes in objective performance among post-call physicians [73,74], fatigued physicians [75], or sleep-deprived physicians [76–78], and Lehmann et al. reported improved performance in post-call physicians [51].

As shown in Supplemental table 4, five of the 26 studies reported conflicting results within their own results due to different cognitive domains that were tested [79–83]. Table 1 provides an overview of the categorisation of all tests that examined the effects of fatigue and sleep deprivation on performance, as well as the nature of the effect (i.e., positive, negative, or no effect). “Attention and concentration” and “memory” (including working memory) are the most commonly assessed domains. With respect to studies that used a test to assess attention and concentration, ten studies found impaired performance, thirteen found no effect, and two studies found improved performance. With respect to memory, seven studies found impaired performance, eight studies found no effect, and two studies found improved performance.

The applicability of the neurocognitive tests used in the included studies was evaluated based on the following characteristics: relevance to the medical profession and the subject of interest; the validity of the tests with respect to measuring impaired performance; the level of the effect of practise; the diversity of the outcomes; the manner in which the examination was conducted; and the duration of the test. Table 2 provides a summary of all of the tests encountered in this review and their

Table 1
Overview of neurocognitive assessment tools and domains of interest (✓, meets requirements; -, does not meet requirements) and performance outcome included studies (↓, negative correlation between test outcomes and contrast of performance; –, no correlation; ↑, positive correlation between test outcomes and contrast of performance).

Test	Visuospatial perception and constructional skills	Executive functioning	Attention and concentration	Mood, personality and behaviour	Memory	Speech, language and communication skills	Author	Outcome
Choice Reaction Time test [127]	–	–	✓	–	–	–	Lee 2003 [70]	––
							Wesnes 1997 [61]	––
Matching-to-sample test [128,129]	✓	–	✓	–	✓	–	O'brien 2012 [71]	––
Psychomotor Vigilance task	–	–	✓	–	–	–	Anderson 2012 [67]	↓
							Saxena 2005 [69]	↓
Spatial Processing Test time	✓	–	–	–	✓	–	O'brien 2012 [71]	↓
Divided Attention task [130]	–	–	✓	–	–	–	Coburn 2006 [97]	––
							Richardson 1996 [75]	––
Grammatical Reasoning Test	–	✓	✓	–	–	–	Deaconson 1988 [76]	––
King–Devick Test	–	–	✓	–	–	✓	Davies 2012 [68]	––
Mathematical processing test [129,131]	–	–	✓	–	✓	–	O'brien 2012[71]	––
Memory Scanning Task	–	–	–	–	✓	–	Wesnes 1997 [61]	↓
Paired Associates Learning [132]	–	–	–	–	✓	–	Veddeng 2014 [49]	↑
							Husby 2014 [81]	↑
Purdue Pegboard Test [133]	–	–	–	–	–	–	Reznick 1987 [78]	––
							Ayalon 2008[66]	––
Stockings of Cambridge	–	✓	–	–	✓	–	Veddeng 2014 [49]	––
							Husby 2014 [81]	––
Connor's CPT-II	–	–	✓	–	–	–	Cavallo 2003 [73]	––
Grooved Pegboard Test [133]	–	–	–	–	–	–	Halbach 2003 [79]	↑
Hopkins Verbal Learning Test [134]	–	–	–	–	✓	–	Chang 2013 [63]	↓
Motor Screening Task (MOT)	–	–	–	–	–	–	Veddeng 2014 [49]	––
							Husby 2014 [81]	––
Pupillographic Sleepiness Test	–	–	✓	–	–	–	Reimann 2009 [82]	↓
Reaction time (RTI) [133]	–	–	✓	–	–	–	Veddeng 2014 [49]	↓
							Husby 2014 [81]	↓
Running Memory scan	–	–	✓	–	✓	–	O'brien 2012 [71]	↓
Simple Reaction Time [133]	–	–	✓	–	–	–	Wesnes 1997 [61]	↓
Trail Making Test [135]	–	✓	✓	–	–	–	Hawkins 1985 [72]	↓
							Deaconson 1988 [76]	––
							Lehmann 2014 [51]	↑
Digit Vigilance test [133,136]	–	–	✓	–	–	–	Wesnes 1997 [61]	↓
Dual N-back Test [137]	–	–	–	–	✓	–	Chang 2013 [63]	––
PASAT [138]	–	–	✓	–	✓	–	Leighton 1983 [60]	↓
							Deaconson 1988 [76]	––
							Reimann 2009 [82]	––
Wisconsin Card Sorting Test [133]	–	✓	–	–	–	–	Khazaie 2010 [77]	––
Picture Recognition	–	–	–	–	✓	–	Wesnes 1997 [61]	↓
CVLT-II [139]	–	–	–	–	✓	–	Halbach 2003 [79]	↓
Iowa Gambling Task	–	✓	–	–	–	–	Khazaie 2010 [77]	––
KAIT [140]	–	✓	–	–	–	–	Dula 2001 [64]	↓
Time Perception Task [141]	–	–	✓	–	✓	–	Khazaie 2010 [77]	––
Minnesota Paper form board test [76]	✓	–	–	–	–	–	Deaconson 1988 [76]	↓
Raven Progressive Matrices [133,142]	–	✓	–	–	–	–	Hawkins 1985 [72]	↓
D2 attention	–	–	✓	–	–	–	Lehmann 2014 [51]	↑

applicability with respect to assessing the physician's performance in a clinical setting. Only four tests met all criteria for applicability in clinical setting: The choice reaction time test (attention and concentration), match-to-sample test (visuospatial perception and constructional skills, attention and concentration, and memory), psychomotor vigilance task (attention and concentration) and the spatial processing test (visuospatial perception and constructional skills, and memory).

Evaluation of sleepiness, sleep loss, and sleep disorders

Although polysomnography has long been the gold standard for objectively quantifying sleep, it is considered somewhat impractical for use in a clinical setting. Sleep statistics are generally obtained more easily using actigraphy to continuously monitor activity; the data can then be linked to subjective perceptions or processed further in order to calculate the risks associated with daytime sleepiness or to make predictions regarding performance and safety.

In total, three studies were eligible for our analysis (Supplemental table 5). Using the outcomes based on the Multiple Sleep Latency Test (MSLT), Reddy et al. reported severe daytime sleepiness among post-call physicians [84]. Surani et al. reported altered MSLT values in the majority of residents studied, but found a paradoxical improvement among post-call residents [85]. Lastly, McCormick et al. processed actigraphy data in a biomathematical model (Sleep Activity Fatigue and Task Effectiveness (SAFTE)) and found prevalent and pervasive levels of fatigue-driven impairment in orthopaedic night-float residents. The calculated predicted overall risk of error, based on an estimated mental effectiveness score (e.g., no direct measurement of medical error or occupational injury), was increased by 22% compared to well-rested historical control data [86].

Table 2
Overview of neurocognitive test characteristics (✓, meets requirements; -, does not meet requirements).

Test	Motor/Cognitive	Relevance	Validation	Practice effects	Multi-domain	Self-test	Duration<10 min
Choice Reaction Time test	Both	✓	✓	✓	✓	✓	✓
Match-to-sample test	Cognitive	✓	✓	✓	✓	✓	✓
Psychomotor Vigilance task	Both	✓	✓	✓	✓	✓	✓
Spatial Processing Test time	Cognitive	✓	✓	✓	✓	✓	✓
Divided Attention task	Cognitive	✓	✓	–	✓	✓	–
Grammatical Reasoning Test	Cognitive	✓	✓	✓	–	✓	✓
King-Devick Test	Both	–	✓	✓	✓	✓	✓
Mathematical processing test	Cognitive	✓	✓	–	✓	✓	
Memory Scanning Task	Cognitive	✓	✓	–	✓	✓	✓
Paired Associates Learning	Cognitive	✓	✓	–	✓	✓	✓
Purdue Pegboard Test	Motor	✓	✓	✓	✓	–	✓
Stockings of Cambridge	Cognitive	✓	✓	✓	–	✓	✓
Connor's CPT-II	Cognitive	–	✓	✓	✓	✓	–
Grooved Pegboard Test	Motor	✓	✓	–	–	✓	✓
Hopkins Verbal Learning Test	Cognitive	–	✓	✓	✓	–	✓
Motor Screening Task (MOT)	Motor	✓	✓	–	–	✓	✓
Pupillographic Sleepiness Test	Cognitive	✓	✓	✓	✓	–	–
Reaction time (RTI)	Cognitive	✓	✓	–	–	✓	✓
Running Memory scan	Cognitive	✓	✓	✓	✓	–	–
Secondary Task Probing	Cognitive	✓	✓	✓	✓	–	–
Simple Reaction Time	Both	✓	✓	–	–	✓	✓
Trail Making Test	Cognitive	–	✓	–	✓	✓	✓
Digit Vigilance test	Both	✓	✓	✓	–	–	–
Dual N-back Test	Cognitive	–	✓	–	–	✓	–
PASAT	Cognitive	–	✓	–	✓	–	✓
Pursuit tracking task	Motor	✓	✓	–	–	✓	–
Wisconsin Card Sorting Test	Cognitive	–	✓	–	–	✓	–
Word and Picture Recognition	Cognitive	–	✓	✓	–	✓	–
CVLT-II	Cognitive	–	✓	✓	–	–	–
Iowa Gambling Task	Cognitive	–	✓	–	–	✓	–
KAIT	Cognitive	–	✓	–	✓	–	–
Time Perception Task	Cognitive	–	✓	✓	–	–	–
Minnesota Paper form board test	Cognitive	–	–	–	✓	–	–
Raven Progressive Matrices	Cognitive	–	✓	–	–	–	–

A total of seven studies performed clinical and/or non-clinical assessments in order to quantify sleep and register changes in the sleep-wake pattern using actigraphy [50,58,77,87,88], polysomnography [89], or both [67]. Overall, all seven studies reported that physicians have significantly less sleep during an on-call period; two of the seven studies [50,77] reported no effect on outcome, whereas the other five studies reported an association between less sleep and poorer outcome.

Evaluation of self-perceived sleepiness, sleep loss, and sleep disorders

Sleepiness, sleep loss, and sleep disorders have also been quantified subjectively using a variety of validated sleep questionnaires, including the Epworth Sleepiness Scale (ESS), the Pittsburgh Sleep Quality Index (PSQI), the Insomnia Severity Index (ISI), the Karolinska Sleepiness Scale (KSS), the Stanford Sleepiness Scale (SSS), the Sleep Deprivation Impact (SDI) scale, and the Visual Analogue Scale (VAS). Eight of the studies included in our review exclusively used subjective tools to assess the effects of sleepiness or sleep dysfunction (Supplemental table 6). Overall, all eight studies described elevated levels of self-reported sleepiness, ranging from moderate to excessive fatigue. In one study, the authors hypothesised that some residents begin their residency with a sleep dysfunction [90]. One study compared ESS to SDI results and found that surgical residents scored as “excessively sleepy” on ESS (scores ranging from 11 to 24), but had lower SDI scores compared to other residents [91]. Two studies examined the association between self-reported sleepiness and self-reported errors and found that higher levels of self-reported sleepiness were independently associated with medical errors [92,93]. One study reported that physicians with higher sleepiness scores performed worse when answering licensing exam questions [94]. Another study reported that excessive sleepiness did not depend on the type of shift, but found that this effect could be overcome by reducing the number of on-call shifts and/or ensuring that residents sleep at least 2 hours per on-call shift [95]. Lastly, two studies examined recovery after considerable sleep loss due to working on-call through the night, but found conflicting results with respect to the adequacy of recovery sleep [87,88].

Sleep questionnaires were also used as an additional assessment tool in seventeen studies. Two of these studies were in a clinical setting [33,96], eight assessed neurocognitive functioning [62,63,68,69,74,81,94,97], four were simulator studies [49,50,52,54], and three used a questionnaire to assess sleep-based performance [84,85,88]. In total, 22 sleep questionnaires were used (Supplemental table 7). In the majority of these studies (71%), higher levels of fatigue were associated with poorer outcome; the remaining studies found no correlation between fatigue and outcome.

Discussion

Our analysis revealed that although a wide range of methods and tools have been used to assess fatigue-related performance impairment among healthcare professionals, the results vary widely and include impairment, no effect, and even improvement of performance. These results should be interpreted with caution as virtually all tools in the summarized studies do not comply with the criteria as formulated by West et al. [18]. Furthermore, three main factors likely contributed to the discrepancy in findings.

The first factor that may have contributed to the conflicting results obtained using neuropsychological tools to evaluate the

effects of fatigue on cognitive performance is the way in which fatigue is defined. In our analysis, we attempted to sort all of the studies into a set of fatigue categories. However, some studies fit better within a specific category than others, possibly explaining the conflicting results found within categories. However, the extent to which physicians were fatigued differed considerably between categories, as “fatigued” describes the difference between the beginning of the day and the end of the day, whereas “post-call” describes the difference between a post-night call and an off-call day. These differences may have affected the study outcomes. Therefore, a better definition of fatigue to differentiate between acute and chronic sleep loss, sleep inertia, and circadian misalignment would be helpful in future studies. The current studies generally lacked the information to provide such a categorization in a reliable fashion.

The second factor that influences the study outcomes is the difference in the amount of rest in the control groups. One cannot assume that a person who is not on call is well rested. Therefore, information about the control state of sleepiness is imperative. One such measure would be to implement the recently developed and validated sleep regularity index (SRI) that is able to not only show sleep duration but also sleep patterns over a longer time [98].

The final factor that was identified to influence outcome is the method used to assess the effect of mental or physical exertion. For example, studies that assessed clinical outcome generally found that sleepiness had no effect on performance. In contrast, simulation studies and studies that use neurocognitive tools found an overall impaired effect, whereas some studies found improved performance.

Each of the assessment methods has both advantages and disadvantages. Using clinical outcome to evaluate the effect of sleepiness or weariness on performance is more reliable when each physician serves as his/her own control, which was the case in five surgical studies [4,15,31–33] and three non-surgical studies [12,39,40]. If results are unmatched, it can be extremely difficult to directly measure the association between sleepiness and clinical outcome. Moreover, in order to assess the effect of sleepiness on clinical performance, an additional detailed objective test for quantifying individual performance impairment is required, thereby providing valuable insights regarding impaired performance due to fatigue and the effect on patient safety; only one study has provided such a measure [22]. Finally even with such a method in place the fact that medical errors are often the result of a group process or a system failure should also be taken into account.

Simulators are commonly used to assess psychomotor skill and have been used recently to provide a surrogate measure of clinical performance. However, it is likely that the motivation and focus of a healthcare professional are lower when using a simulator compared to performing a real-life procedure, and this might have a negative effect on performance, thereby affecting the study outcome [88]. Therefore, whether the results obtained using a simulator are a suitable surrogate for clinical performance is debatable. Nonetheless, they do provide important insight into impaired cognitive and/or motor functions, and as such could serve as a suitable measure of specific technical skills and certain objectified clinical procedures.

Actigraphy is generally believed to be suitable for assessing long-term sleep patterns. Thus, this non-invasive instrument was used to monitor sleep and sleep patterns either as a primary ($n = 1$ study) or secondary ($n = 7$ studies) outcome. However, actigraphy uses an indirect measure (physical movement) to quantify the physiological state of sleep, rather than using a direct measure of sleep quality [99,100].

It has been argued that actigraphy has a low ability to detect wakefulness bouts within sleep periods. Given the nature of the medical profession (e.g., shift work and fragmented sleep), this might not be the most suitable method for assessing sleep quality and correlate outcomes to clinical performance [101]. Due to these limitations it has been advised to use complementary tests (both subjective and objective) in order to obtain more detailed sleep-wake information [102]. Alternatively, polysomnography provides a direct measure of sleep, based on several physiological signals (e.g., brain activity (EEG), eye movements (EOG) and muscle activation (EMG)). Hereby failures of cognitive control or attention can be detected as reported by Lockley et al. in a randomized controlled trial [89], in theory this method would be suitable to evaluate sleep impaired performance. However, in view of the aforementioned practical objections, this method is less suitable as a self-test for performance evaluation.

Moreover, actigraphy data and scheduling information can be processed in mathematical models in order to predict the effect of fatigue or fatigue risk [103]. These models have been used to provide a continuous indication of performance by several agencies and industries, including the US Department of Defense and the aviation and automotive industries [104]. Unfortunately, fatigue optimization scheduling to reduce medical errors has not been explored elaborately up to now. These models are sensitive to individual variability; therefore, relying on sleepiness estimates based on group averages might be problematic due to significant differences in schedules and sleep variability between institutions, individuals and workdays throughout different medical professions [104,105]. Moreover, calculations are based on data collected over a period of time, these models do not directly measure sleepiness or actual real-time feedback, or objective measures of mental fatigue, but data can be used to retrospectively assess the risks of a specific schedule allowing implementation of countermeasures. In this respect it might be more useful to rely on alternative systems such as the unified model of performance [106]. Lastly, up to now these mathematical models have not been elaborately validated within the medical field, and it would be less appropriate to make decisions concerning individuals based on model estimates.

We found that studies using neuropsychological assessments to evaluate the effects of mental exertion on cognitive performance yielded conflicting results. This discrepancy may be due in part to various tests that were used to measure performance and/or the combinations of neurocognitive domains that were tested. Although “memory” and “attention” were the most commonly evaluated domains, no clear outcome emerged. This may mean that sleepiness does not necessarily impair cognitive performance; alternatively – and more likely – some tests may simply not be sensitive enough to detect the effects of mental exertion. Interestingly, some studies found that sleepiness actually improved cognitive performance. This paradoxical finding might be due to inadequately controlling for the learning effect of practise [79–81].

Therefore, although using a neurocognitive tool might provide a clear outcome with respect to impaired performance, some tools might be less relevant to the specific research question, and the respective results should be interpreted with caution. We found that four tests were validated with respect to performance impairment, are relevant for the medical profession, have an acceptable level of the effect of practice, are sensitive to test multiple applicable domains, can be used as self-test and do not last longer than ten minutes. Of these four, only the choice reaction time test and psychomotor vigilance task assess both motor- and cognitive function and would be suitable for use in a clinical setting.

Other studies used self-evaluation – either exclusively or as a supplementary test – to assess performance. Regardless of the study design, self-evaluation is extremely important, as awareness of one's own skills and the ability to self-predict performance is recognised by the US Accreditation Council for Graduate Medical

Education (ACGME) as a core competency [107]. In addition, perceived mental and physical exertion can have a major impact on personal life and on the ability to perform and learn [108,109]. However, some of the questionnaires used in the included studies are not necessarily suitable for real-time assessment. For example, the ISI quantifies the nature and symptoms of sleep problems and is used to measure insomnia [110], the ESS – despite being validated and cost-effective – focusses on habitual sleepiness [91,111], and the PSQI retrospectively evaluates sleep quality during the preceding month [112]. In contrast, the VAS, SDI, SSS, and KSS provide a measure of existing sleepiness and are therefore more appropriate [113]. The VAS, which is a psychometric response scale, has been validated for use in measuring the effects of sleep deprivation [114], alcohol [115], depression [116], and sedatives [117]. The SDI uses a five-point Likert scale to quantify the effect of current sleep deprivation on performance [91]. The SSS uses a progressive scale ranging from high alertness to imminent sleep in order to quantify sleepiness [118]. Finally, the KSS uses a nine-point Likert scale and has high validity for assessing the current level of drowsiness [49,50,81,113].

Compared to other professionals, healthcare professionals are generally more likely to deny the effects of stress and fatigue and subsequently minimise their personal susceptibility to commit an error [119]. Overall, physicians tend to have a limited ability to accurately self-assess their performance [16,111], which can lead to a more favourable outcome. As a result, regardless of the test being used, a physician's self-assessment of performance will be distorted if the physician does not recognise the effects of sleep deprivation or does not acknowledge his/her own stress and/or sleepiness [94]. Similarly, the outcome of a questionnaire can be affected by recall and/or reporting bias. On the other hand, given that sleepiness scores have been correlated to clinical errors, low exam results, stress, reduced job satisfaction, and self-perceived medical errors, they give valuable insight into one's self-perceived mental state and quality of life and may provide an additional domain in the physician's self-assessment [116,120].

As discussed above, clinical assessment is suitable for determining patient safety when the results are matched, but it is not necessarily a practical approach for assessing real-time performance. With respect to simulations, in addition to the aforementioned disadvantages, this approach is not always practical, as it takes considerable time and does not yet fully mimic real-life multidimensional clinical conditions. Actigraphy – specifically, processed sleep quality data – can be used to provide a calculated risk with regard to performance but does not provide direct insight and is likely too time-consuming and therefore impractical for use in a clinical setting. With respect to neurocognitive tests, their clinical relevance is debatable, and extensive tools that assess multiple domains may be too time-consuming to be feasible. Nevertheless, if used to measure the most relevant domains, these tools may be quite useful for assessing performance; in this respect, further research is needed in order to determine which domains are appropriate for assessing clinically relevant cognitive performance in healthcare professionals. Lastly, self-assessment tools should not be used exclusively to measure performance, as many physicians might be unable to accurately assess their own performance.

To overcome these issues, we recommend using a tool with the following characteristics, a short-duration, and real-time objective and subjective measures of performance. The tests used in this tool should be sufficiently sensitive to measure the relevant key domains depending on the objective(s). In addition, most outcomes are reported using an arbitrary scale and therefore may not make sense to readers who are unfamiliar with the methodology. Therefore, a specific frame of reference (i.e., multifactorial validity) is needed in order to clearly illustrate the relationship between study results and clinical endpoints or social relevance. Ideally, study outcomes should be validated using a specific reference standard for evaluating performance [16]. For example, international guidelines for research on drugged driving recommend that the effects of ethanol serve as a reference standard for assessing impaired performance impairment [121,122]. Validation of results using blood alcohol concentration provides a socially accepted – and legal – frame of reference. [123,124] Several studies have followed this recommendation, validating performance with the effects induced by ethanol using subjective and objective neuropsychological tests [62,86,115,123–126].

Conclusions

In this review, we provide an overview of the various methods by which researchers assess fatigue-impaired performance in healthcare professionals. To date, no clear methodology is available to provide support with regard to the relation between impaired performance and reduced patient safety, and objective data regarding performance among healthcare professionals is scarce. Thus, an easy-to-use and well-validated assessment tool is urgently needed. In addition to providing the much-needed missing data, such a tool will also facilitate the development of guidelines regarding the combination of shift duration and work circumstances that will maximise safe patient care. This is also true for proposed measures designed to reduce fatigue. Until objective data are available, caution is warranted when introducing regulations that can have significant implications with respect to physicians, education, scheduling, and – ultimately – patient care.

Practice points

- Objective comprehensive measurement methods regarding performance among physically or mentally exerted healthcare professionals are currently unavailable;
- Previously published contrasting outcomes can therefore be attributed to three main factors:
  a. the wide variety in the methods used;
  b. differences in the operationalisation of fatigue;
  c. study design with multiple poorly defined covariates.
- Until objective comprehensive data become available, caution should be used when developing regulations that can have implications for physicians, education, manpower planning, and – ultimately – patient care.

Research Agenda

- The implementation of a clinically applicable tool that can provide uniform data regarding fatigue-impaired performance.
- Initiation of large well-controlled trials to provide this data.
- Uniform data regarding fatigue-impaired performance should be related to detailed individual outcomes of patient care.

Conflicts of interest

The authors declare that there is no conflict of interest.

Acknowledgements

The authors thank professor J.H. Meijer, for insightful comments and critical proof reading. This research was funded by grants from the Scientific Committee of the Haaglanden Medical Center and the Stichting Kwaliteitsgelden Medisch Specialisten (SKMS).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.smrv.2019.04.007.

References

*[1] Landrigan CP, Rothschild JM, Cronin JW, Kaushal R, Burdick E, Katz JT, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. N Engl J Med 2004;351:1838–48. https://doi.org/10.1056/NEJMoa041406.

[2] Thomas M, Allen MS, Wigle DA, Shen KR, Cassivi SD, Nichols III FC, et al. Does surgeon workload per day affect outcomes after pulmonary lobectomies? Ann Thorac Surg 2012;94:966–72. https://doi.org/10.1016/j.athoracsur.2012.04.099.

*[3] Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard medical practice study I. N Engl J Med 1991;324:370–6. https://doi.org/10.1056/NEJM199102073240604.

[4] Lee KT, Mun GH. Is after-hours free-flap surgery associated with adverse outcomes? J Plast Reconstr Aesthet Surg 2013;66:460–6.

[5] West CP, Tan AD, Shanafelt TD. Association of resident fatigue and distress with occupational blood and body fluid exposures and motor vehicle incidents. Mayo Clin Proc 2012;87:1138–44. https://doi.org/10.1016/j.mayocp.2012.07.021.

[6] Veasey S, Rosen R, Barzansky B, Rosen I, Owens J. Sleep loss and fatigue in residency training: a reappraisal. Jama 2002;288:1116–24.

[7] Lee ML, Howard ME, Horrey WJ, Liang Y, Anderson C, Shreeve MS, et al. High risk of near-crash driving events following night-shift work. Proc Natl Acad Sci USA 2016;113:176–81. https://doi.org/10.1073/pnas.1510383112.

*[8] Nasca TJ, Day SH, Amis Jr ES. The new recommendations on duty hours from the ACGME Task Force. N Engl J Med 2010;363:e3.

*[9] The Lancet. Doctors' training and the European working time directive. Lancet 2010;375:2121. https://doi.org/10.1016/S0140-6736(10)60977-4.

[10] Pellegrini CA, Britt LD, Hoyt DB. Sleep deprivation and elective surgery. N Engl J Med 2010;363:2672–3. https://doi.org/10.1056/NEJMc1012517.

*[11] Ahmed N, Devitt KS, Keshet I, Spicer J, Imrie K, Feldman L, et al. A systematic review of the effects of resident duty hour restrictions in surgery: impact on resident wellness, training, and patient outcomes. Ann Surg 2014;259:1041–53. https://doi.org/10.1097/SLA.0000000000000595.

[12] Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Kim J. Long radiology workdays reduce detection and accommodation accuracy. J Am Coll Radiol 2010;7:698–704. https://doi.org/10.1016/j.jacr.2010.03.004.

*[13] Nurok M, Czeisler CA, Lehmann LS. Sleep deprivation, elective surgical procedures, and informed consent. N Engl J Med 2010;363:2577–9. https://doi.org/10.1056/NEJMp1007901.

[14] Hendey GW, Barth BE, Soliz T. Overnight and postcall errors in medication orders. Acad Emerg Med 2005;12:629–34. https://doi.org/10.1197/j.aem.2005.02.009.

[15] Bagrodia A, Rachakonda V, Delafuente K, Toombs S, Yeh O, Scales J, et al. Surgeon fatigue: impact of case order on perioperative parameters and patient outcomes. J Urol 2012;188:1291–6. https://doi.org/10.1016/j.juro.2012.06.021.

[16] Huizinga CRH, de Kam ML, Stockmann HBAC, van Gerven JMA, Cohen AF, van der Bogt KEA. Evaluating fitness to perform in surgical residents after night shifts and alcohol intoxication: the development of a "Fit-to-Perform" test. J Surg Educ 2018. https://doi.org/10.1016/j.jsurg.2018.01.010.

[17] Orman ES, Hayashi PH, Dellon ES, Gerber DA, Barritt AS. Impact of nighttime and weekend liver transplants on graft and patient outcomes. Liver Transplant 2012;18:558–65. https://doi.org/10.1002/lt.23395.

[18] West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, et al. Systems to rate the strength of scientific evidence. Evid Rep Technol Assess (Summ) 2002:1–11.

[19] Marshall J. The handbook of clinical neuropsychology. OUP Oxford; 2012.

[20] Peskun C, Walmsley D, Waddell J, Schemitsch E. Effect of surgeon fatigue on hip and knee arthroplasty. Can J Surg 2012;55:81–6. https://doi.org/10.1503/cjs.032910.

[21] Fechner G, Pezold C, Hauser S, Gerhardt T, Muller SC. Kidney"s nightshift, kidney"s nightmare? Comparison of daylight and nighttime kidney transplantation: impact on complications and graft survival. Transplant Proc 2008;40:1341–4. https://doi.org/10.1016/j.transproceed.2008.02.072.

[22] de Boer J, Van der Bogt K, Putter H, Ooms-de Vries K, Haase-Kromwijk B, Pol R, et al. Surgical quality in organ procurement during day and night: an analysis of quality forms. BMJ Open 2018;8:e022182. https://doi.org/10.1136/bmjopen-2018-022182.

[23] Shaw TM, Lonze BE, Feyssa EL, Segev DL, May N, Parsikia A, et al. Operative start times and complications after kidney transplantation. Clin Transplant 2012;26:E177–83.

[24] Yaghoubian A, Kaji AH, Putnam B, de Virgilio C. Trauma surgery performed by "sleep deprived" residents: are outcomes affected? J Surg Educ 2010;67:449–51. https://doi.org/10.1016/j.jsurg.2010.09.003.

[25] Yaghoubian A, Kaji AH, Ishaque B, Park J, Rosing DK, Lee S, et al. Acute care surgery performed by sleep deprived residents: are outcomes affected? J Surg Res 2010;163:192–6.

[26] Sharpe JP, Weinberg JA, Magnotti LJ, Nouer SS, Yoo W, Zarzaur BL, et al. Outcomes of operations performed by attending surgeons after overnight trauma shifts. J Am Coll Surg 2013;216. https://doi.org/10.1016/j.jamcollsurg.2012.12.005. 791–7–discussion797–9.

[27] Schieman C, MacLean AR, Buie WD, Rudmik LR, Ghali WA, Dixon E. Does surgeon fatigue influence outcomes after anterior resection for rectal cancer? Am J Surg 2008;195. https://doi.org/10.1016/j.amjsurg.2008.01.009. 684–7–discussion687–8.

[28] Haynes DF, Schwedler M, Dyslin DC, Rice JC, Kerstein MD. Are postoperative complications related to resident sleep deprivation? South Med J 1995;88:283–9.

[29] Ellman PI, Kron IL, Alvis JS, Tache-Leon C, Maxey TS, Reece TB, et al. Acute sleep deprivation in the thoracic surgical resident does not affect operative outcomes. Ann Thorac Surg 2005;80. https://doi.org/10.1016/j.athoracsur.2005.01.034. 60–4–discussion64–5.

*[30] Ellman PI, Law MG, Tache-Leon C, Reece TB, Maxey TS, Peeler BB, et al. Sleep deprivation does not affect operative results in cardiac surgery. Ann Thorac Surg 2004;78. https://doi.org/10.1016/j.athoracsur.2004.04.006. 906–11–discussion906–11.

*[31] Vinden C, Nash DM, Rangrej J, Shariff SZ, Dixon SN, Jain AK, et al. Complications of daytime elective laparoscopic cholecystectomies performed by surgeons who operated the night before. Jama 2013;310:1837–41. https://doi.org/10.1001/jama.2013.280372.

[32] Rothschild JM. Risks of complications by attending physicians after performing nighttime procedures. Jama 2009;302:1565–72. https://doi.org/10.1001/jama.2009.1423.

[33] Hayter MA, Friedman Z, Katznelson R, Hanlon JG, Borges B, Naik VN. Effect of sleep deprivation on labour epidural catheter placement. Br J Anaesth 2010;104:619–27. https://doi.org/10.1093/bja/aeq063.

[34] Ben-Aroya Z, Segal D, Hadar A, Hallak M, Friger M, Katz M, et al. Effect of OB/GYN residents' fatigue and training level on the accuracy of fetal weight estimation. Fetal Diagn Ther 2002;17:177–81. https://doi.org/10.1007/s00464-014-3503-0.

[35] Ruutiainen AT, Durand DJ, Scanlon MH, Itri JN. Increased error rates in preliminary reports issued by radiology residents working more than 10 consecutive hours overnight. Acad Radiol 2013;20:305–11.

[36] Davenport MS, Ellis JH, Khalatbari SH, Myles JD, Klein KA. Effect of work hours, caseload, shift type, and experience on resident call performance. Acad Radiol 2010;17:921–7. https://doi.org/10.1016/j.acra.2010.03.006.

[37] Terreblanche OD, Andronikou S, Hlabangana LT, Brown T, Boshoff PE. Should registrars be reporting after-hours CT scans? A calculation of error rate and the influencing factors in South Africa. Acta Radiol 2012;53:61–8. https://doi.org/10.1258/ar.2011.110103.

[38] Davydov L, Caliendo G, Mehl B, Smith LG. Investigation of correlation between house-staff work hours and prescribing errors. Am J Health Syst Pharm 2004;61:1130–4.

[39] Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Madsen MT, Kramer DJ. Do long radiology workdays affect nodule detection in dynamic CT interpretation? J Am Coll Radiol 2012;9:191–8. https://doi.org/10.1016/j.jacr.2011.11.013.

[40] Christensen EE, Dietz GW, Murry RC, Moore JG. The effect of fatigue on resident performance. Radiology 1977;125:103–5. https://doi.org/10.1016/j.jsurg.2010.09.003.

[41] Amirian I, Mortensen JF, Rosenberg J, Gögenur I. Admission medical records made at night time have the same quality as day and evening time records. Dan Med J 2014;61:A4868.

[42] Gates M, Wingert A, Featherstone R, Samuels C, Simon C, Dyson MP. Impact of fatigue and insufficient sleep on physician and patient outcomes: a systematic review. BMJ Open 2018;8:e021967. https://doi.org/10.1136/bmjopen-2018-021967.

[43] Taffinder NJ, McManus IC, Gul Y, Russell RC, Darzi A. Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. Lancet 1998;352:1191.

* The most important references are denoted by an asterisk.

[44] Tsafrir Z, Korianski J, Almog B, Many A, Wiesel O, Levin I. Effects of fatigue on residents' performance in laparoscopy. J Am Coll Surg 2015;221:564–70. https://doi.org/10.1016/j.jamcollsurg.2015.02.024. e3.

[45] Kahol K, Leyba MJ, Deka M, Deka V, Mayes S, Smith M, et al. Effect of fatigue on psychomotor and cognitive skills. Am J Surg 2008;195:195–204. https://doi.org/10.1016/j.amjsurg.2007.10.004.

[46] Eastridge BJ, Hamilton EC, O'Keefe GE, Rege RV, Valentine RJ, Jones DJ, et al. Effect of sleep deprivation on the performance of simulated laparoscopic surgical skill. Am J Surg 2003;186:169–74. https://doi.org/10.1016/S0002-9610(03)00183-1.

[47] Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Laparoscopic performance after one night on call in a surgical department: prospective study. Bmj 2001;323:1222–3.

[48] Gerdes J, Kahol K, Smith M, Leyba MJ, Ferrara JJ. Jack Barney award: the effect of fatigue on cognitive and psychomotor skills of trauma residents and attending surgeons. Am J Surg 2008;196. https://doi.org/10.1016/j.amjsurg.2008.07.030. 813–9–discussion819–20.

[49] Veddeng A, Husby T, Engelsen IB, Kent A, Flaatten H. Impact of night shifts on laparoscopic skills and cognitive function among gynecologists. Acta Obstet Gynecol Scand 2014;93:1255–61. https://doi.org/10.1111/aogs.12496.

[50] Amirian I, Andersen LT, Rosenberg J, Gögenur I. Laparoscopic skills and cognitive function are not affected in surgeons during a night shift. J Surg Educ 2014;71:543–50. https://doi.org/10.1016/j.jsurg.2013.12.007.

[51] Lehmann KS, Martus P, Little-Elk S, Maass H, Holmer C, Zurbuchen U, et al. Impact of sleep deprivation on medium-term psychomotor and cognitive performance of surgeons: prospective cross-over study with a virtual surgery simulator and psychometric tests. Surgery 2010;147:246–54. https://doi.org/10.1016/j.surg.2009.08.007.

[52] Uchal M, Tjugum J, Martinsen E, Qiu X, Bergamaschi R. The impact of sleep deprivation on product quality and procedure effectiveness in a laparoscopic physical simulator: a randomized controlled trial. Am J Surg 2005;189:753–7. https://doi.org/10.1016/j.amjsurg.2005.03.021.

[53] DeMaria EJ, McBride CL, Broderick TJ, Kaplan BJ. Night call does not impair learning of laparoscopic skills. Surg Innovat 2005;12:145–9.

[54] Schlosser K, Maschuw K, Kupietz E, Weyers P, Schneider R, Rothmund M, et al. Call-associated acute fatigue in surgical residents–subjective perception or objective fact? A cross-sectional observational study to examine the influence of fatigue on surgical performance. World J Surg 2012;36:2276–87. https://doi.org/10.1007/s00268-012-1699-5.

[55] Brandenberger J, Kahol K, Feinstein AJ, Ashby A, Smith M, Ferrara JJ. Effects of duty hours and time of day on surgery resident proficiency. Am J Surg 2010;200. https://doi.org/10.1016/j.amjsurg.2010.06.009. 814–8–discussion818–9.

[56] Olasky J, Chellali A, Sankaranarayanan G, Zhang L, Miller A, De S, et al. Effects of sleep hours and fatigue on performance in laparoscopic surgery simulators. Surg Endosc 2014. https://doi.org/10.1007/s00464-014-3503-0.

[57] Gordon JA, Alexander EK, Lockley SW, Flynn-Evans E, Venkatan SK, Landrigan CP, et al. Does simulator-based clinical performance correlate with actual hospital behavior? The effect of extended work hours on patient care provided by medical interns. Acad Med 2010;85:1583–8. https://doi.org/10.1097/ACM.0b013e3181f073f0.

[58] Murray D, Dodds C. The effect of sleep disruption on performance of anaesthetists–a pilot study. Anaesthesia 2003;58:520–5.

[59] Storer JS, Floyd HH, Gill WL, Wimbish GC, Ginsberg H. Effects of sleep deprivation of cognitive ability and skills of pediatrics residents. Acad Med 1989;64:29–32.

[60] Leighton K, Livingston M. Fatigue in doctors. Lancet 1983;1:1280.

[61] Wesnes KA, Walker MB, Walker LG, Heys SD, White L, Warren R, et al. Cognitive performance and mood after a weekend on call in a surgical unit. Br J Surg 1997;84:493–5.

[62] Bartel P, Offermeier W, Smith F, Becker P. Attention and working memory in resident anaesthetists after night duty: group and individual effects. Occup Environ Med 2004;61:167–70.

[63] Chang LC, Mahoney JJ, Raty SR, Ortiz J, Apodaca S, La Garza De R. Neurocognitive effects following an overnight call shift on faculty anesthesiologists. Acta Anaesthesiol Scand 2013;57:1051–7. https://doi.org/10.1111/aas.12120.

[64] Dula DJ, Dula NL, Hamrick C, Wood GC. The effect of working serial night shifts on the cognitive functioning of emergency physicians. Ann Emerg Med 2001;38:152–5. https://doi.org/10.1067/mem.2001.116024.

[65] Flinn F, Armstrong C. Junior doctors' extended work hours and the effects on their performance: the Irish case. Int J Qual Health Care 2011;23:210–7. https://doi.org/10.1093/intqhc/mzq088.

[66] Ayalon RD, Friedman Jr F. The effect of sleep deprivation on fine motor coordination in obstetrics and gynecology residents. Am J Obstet Gynecol 2008;199:576. https://doi.org/10.1016/j.ajog.2008.06.080. e1–5.

[67] Anderson C, Sullivan JP, Flynn-Evans EE, Cade BE, Czeisler CA, Lockley SW. Deterioration of neurobehavioral performance in resident physicians during repeated exposure to extended duration work shifts. Sleep 2012;35:1137–46. https://doi.org/10.5665/sleep.2004.

[68] Davies EC, Henderson S, Balcer LJ, Galetta SL. Residency training: the King-Devick test and sleep deprivation: study in pre- and post-call neurology residents. Neurology 2012;78:e103–6. https://doi.org/10.1212/WNL.0b013e318251833d.

[69] Saxena AD, George CFP. Sleep and motor performance in on-call internal medicine residents. Sleep 2005;28:1386–91.

[70] Lee DT, Chan SW, Kwok SP. Introduction of night shift call system for surgical trainees: a prospective self-controlled trial. Med Educ 2003;37:495–9.

[71] O'Brien MJ, O'Toole RV, Newell MZ, Lydecker AD, Nascone J, Sciadini M, et al. Does sleep deprivation impair orthopaedic surgeons' cognitive and psychomotor performance? J Bone Joint Surg Am 2012;94:1975–81. https://doi.org/10.2106/JBJS.K.00958.

[72] Hawkins MR, Vichick DA, Silsby HD, Kruzich DJ, Butler R. Sleep and nutritional deprivation and performance of house officers. J Med Educ 1985;60:530–5. https://doi.org/10.1308/003588412X13171221589018.

[73] Cavallo A, Ris MD, Succop P. The night float paradigm to decrease sleep deprivation: good solution or a new problem? Ergonomics 2003;46:653–63. https://doi.org/10.1080/0014013031000085671.

[74] Cuthbertson DW, Bershad EM, Sangi Haghpeykar H, Cohen HS. Balance as a measurement of fatigue in postcall residents. Laryngoscope 2015;125:337–41. https://doi.org/10.1002/lary.24792.

[75] Richardson GS, Wyatt JK, Sullivan JP, Orav EJ, Ward AE, Wolf MA, et al. Objective assessment of sleep and alertness in medical house staff and the impact of protected time for sleep. Sleep 1996;19:718–26.

[76] Deaconson TF, O'Hair DP, Levy MF, Lee MB, Schueneman AL, Codon RE. Sleep deprivation and resident performance. Jama 1988;260:1721–7.

[77] Khazaie H, Tahmasian M, Ghadami MR, Safaei H, Ekhtiari H, Samadzadeh S, et al. The effects of chronic partial sleep deprivation on cognitive functions of medical residents. Iran J Psychiatry 2010;5:74–7. https://doi.org/10.1016/j.jsurg.2012.04.005.

[78] Reznick RK, Folse JR. Effect of sleep deprivation on the performance of surgical residents. Am J Surg 1987;154:520–5.

[79] Halbach MM, Spann CO, Egan G. Effect of sleep deprivation on medical resident and student cognitive function: a prospective study. Am J Obstet Gynecol 2003;188:1198–201.

[80] Rubin R, Orris P, Lau SL, Hryhorczuk DO, Furner S, Letz R. Neurobehavioral effects of the on-call experience in housestaff physicians. J Occup Med 1991;33:13–8.

[81] Husby T, Torgersen J, Flaatten H. Cognitive effects of hospital calls in anaesthesiologists. Acta Anaesthesiol Scand 2014;58:177–84. https://doi.org/10.1111/aas.12220.

[82] Reimann M, Manz R, Prieur S, Reichmann H, Ziemssen T. Education research: cognitive performance is preserved in sleep-deprived neurology residents. Neurology 2009;73:e99–103.

[83] Cao CG, Weinger MB, Slagle J, Zhou C, Ou J, Gillin S, et al. Differences in day and night shift clinical performance in anesthesiology. Hum Factors 2008;50:276–90.

[84] Reddy R, Guntupalli K, Alapat P, Surani S, Casturi L, Subramanian S. Sleepiness in medical ICU residents. Chest 2009;135:81–5.

[85] Surani S, Subramanian S, Aguillar R, Ahmed M, Varon J. Sleepiness in medical residents: impact of mandated reduction in work hours. Sleep Med 2007;8:90–3. https://doi.org/10.1016/j.sleep.2006.08.018.

[86] McCormick F, Kadzielski J, Landrigan CP, Evans B, Herndon JH, Rubash HE. Surgeon fatigue: a prospective analysis of the incidence, risk, and intervals of predicted fatigue-related impairment in residents. Arch Surg 2012;147:430–5. https://doi.org/10.1001/archsurg.2012.84.

[87] Malmberg B, Kecklund G, Karlson B, Persson R, Flisberg P, Ørbaek P. Sleep and recovery in physicians on night call: a longitudinal field study. BMC Health Serv Res 2010;10:239. https://doi.org/10.1186/1472-6963-10-239.

[88] Rose M, Manser T, Ware JC. Effects of call on sleep and mood in internal medicine residents. Behav Sleep Med 2008;6:75–88. https://doi.org/10.1080/15402000801952914.

*[89] Lockley SW, Cronin JW. Effect of reducing interns' weekly work hours on sleep and attentional failures. N Engl J Med 2004;351:1829–37.

[90] Pulliam SJ, Weinstein DF, Malhotra A, Macklin EA, Berkowitz LR. Baseline sleep dysfunction among matriculating interns. J Grad Med Educ 2012;4:202–8.

[91] Woodrow SI, Park J, Murray BJ, Wang C, Bernstein M, Reznick RK, et al. Differences in the perceived impact of sleep deprivation among surgical and non-surgical residents. Med Educ 2008;42:459–67. https://doi.org/10.1111/j.1365-2923.2007.02963.x.

[92] Gander P, Purnell H, Garden A, Woodward A. Work patterns and fatigue-related risk among junior doctors. Occup Environ Med 2007;64:733–8. https://doi.org/10.1136/oem.2006.030916.

*[93] West CP. Association of resident fatigue and distress with perceived medical errors. Jama 2009;302:1294–300. https://doi.org/10.1001/jama.2009.1389.

[94] Durning SJ, Capaldi VF, Artino AR, Graner J, van der Vleuten C, Beckman TJ, et al. A pilot study exploring the relationship between internists' self-reported sleepiness, performance on multiple-choice exam items and prefrontal cortex activity. Med Teach 2014;36:434–40. https://doi.org/10.3109/0142159X.2014.888408.

[95] Hanlon JG, Hayter MA, Bould MD, Joo HS, Naik VN. Perceived sleepiness in Canadian anesthesia residents: a national survey. Can J Anaesth 2009;56:27–34. https://doi.org/10.1007/s12630-008-9003-8.

[96] Hansen KN, Marshall AJ, Heitmann A, Santiago MJ. 212: resident alertness, stress, and self-reported medical errors in an Urban teaching hospital emergency department. Ann Emerg Med 2009;54:S65–6.

[97] Coburn M. Influence of a 30-min break on divided attention and working memory in resident anaesthetists on daily routine. Br J Anaesth 2006;97:315–9. https://doi.org/10.1093/bja/ael151.
[98] Phillips AJK, Clerx WM, O'Brien CS, Sano A, Barger LK, Picard RW, et al. Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. Sci Rep 2017;7:3216. https://doi.org/10.1038/s41598-017-03171-4.
[99] Russell CA, Caldwell JA, Arand D, Myers LJ, Wubbels P, Downs H. Controlling fatigue risk with the Fatigue Science ReadiBand actigraph and associated sleep/wake classification algorithms: use case and validation. 2011.
[100] Russell CA, Caldwell JA, Arand D, Arch LM. Validation of the fatigue science readiband actigraph and associated sleep/wake classification algorithms. Fatiguesciencecom. 2000.
[101] Paquet J, Kawinska A, Carrier J. Wake detection capacity of actigraphy during sleep. Sleep 2007;30:1362–9.
[102] Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med Rev 2011;15:259–67. https://doi.org/10.1016/j.smrv.2010.10.001.
[103] Van Dongen HPA. Comparison of mathematical model predictions to experimental data of fatigue and performance. Aviat Space Environ Med 2004;75:A15–36.
[104] Hursh SR, Redmond DP, Johnson ML, Thorne DR, Belenky G, Balkin TJ, et al. Fatigue models for applied research in warfighting. Aviat Space Environ Med 2004;75. A44–53–discussionA54–60.
[105] McCormick F, Kadzielski J, Evans BT, Landrigan CP, Herndon J, Rubash H. Fatigue optimization scheduling in graduate medical education: reducing fatigue and improving patient safety. J Grad Med Educ 2013;5:107–11. https://doi.org/10.4300/JGME-D-12-00021.1.
[106] Liu J, Ramakrishnan S, Laxminarayan S, Balkin TJ, Reifman J. Real-time individualization of the unified model of performance. J Sleep Res 2017;26:820–31. https://doi.org/10.1111/jsr.12535.
[107] de Blacam C, O'Keeffe DA, Nugent E, Doherty E, Traynor O. Are residents accurate in their assessments of their own surgical skills? Am J Surg 2012;204:724–31. https://doi.org/10.1016/j.amjsurg.2012.03.003.
[108] Papp KK, Stoller EP, Sage P, Aikens JE, Owens J, Avidan A, et al. The effects of sleep loss and fatigue on resident-physicians: a multi-institutional, mixed-method study. Acad Med 2004;79:394–406.
[109] Gravenstein JS, Cooper JB, Orkin FK. Work and rest cycles in anesthesia practice. Anesthesiology 1990;72:737–42.
[110] Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001;2:297–307.
[111] Heaton K, Davis D, Mazmanian P, Anderson D, Fordis M, Van Harrison R, et al. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. Jama 2006;296:1094–102. https://doi.org/10.1001/jama.296.9.1094.
[112] Backhaus J, Junghanns K, Broocks A, Riemann D, Hohagen F. Test-retest reliability and validity of the Pittsburgh sleep quality index in primary insomnia. J Psychosom Res 2002;53:737–40.
[113] Kaida K, Takahashi M, Akerstedt T, Nakata A, Otsuka Y, Haratani T, et al. Validation of the Karolinska sleepiness scale against performance and EEG variables. Clin Neurophysiol 2006;117:1574–81. https://doi.org/10.1016/j.clinph.2006.03.011.
[114] Van Steveninck AL, van Berckel BN, Schoemaker RC, Breimer DD, van Gerven JM, cohen AF. The sensitivity of pharmacodynamic tests for the central nervous system effects of drugs on the effects of sleep deprivation. J Psychopharmacol (Oxford) 1999;13:10–7. https://doi.org/10.1177/026988119901300102.
[115] Zoethout RWM, Schoemaker RC, Zuurman L, van Pelt H, Dahan A, Cohen AF, et al. Central nervous system effects of alcohol at a pseudo-steady-state concentration using alcohol clamping in healthy volunteers. Br J Clin Pharmacol 2009;68:524–34. https://doi.org/10.1111/j.1365-2125.2009.03488.x.
[116] Kertzman S, Aladjem Z, Milo R, Ben-Nahum Z, Birger M, Grinspan H, et al. The utility of the Visual Analogue Scale for the assessment of depressive mood in cognitively impaired patients. Int J Geriatr Psychiatry 2004;19:789–96. https://doi.org/10.1002/gps.1141.
[117] Van Steveninck AL, Gieschke R, Schoemaker HC, Pieters MS, Kroon JM, Breimer DD, et al. Pharmacodynamic interactions of diazepam and intravenous alcohol at pseudo steady state. Psychopharmacology (Berl) 1993;110:471–8. https://doi.org/10.1007/BF02244655.
[118] Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. Psychophysiology 1973;10:431–6.
[119] Sexton JB, Thomas EJ, Helmreich RL. Error, stress, and teamwork in medicine and aviation: cross sectional surveys. Bmj 2000;320:745–9.
[120] Nyenhuis DL, Yamamoto C, Stern RA, Luchetta T, Arruda JE. Standardization and validation of the visual analog mood scales. Clin Neuropsychol 1997;11:407–15. https://doi.org/10.1080/13854049708400470.
[121] Walsh JM, Verstraete AG, Huestis MA, Mørland J. Guidelines for research on drugged driving. Addiction 2008;103:1258–68. https://doi.org/10.1111/j.1360-0443.2008.02277.x.
[122] Jongen S, Vuurman EFPM, Ramaekers JG, Vermeeren A. The sensitivity of laboratory tests assessing driving related skills to dose-related impairment of alcohol: a literature review. Accid Anal Prev 2016;89:31–48. https://doi.org/10.1016/j.aap.2016.01.001.
[123] Dawson D, Reid K. Fatigue, alcohol and performance impairment. Nature 1997;388:235. https://doi.org/10.1038/40775.
[124] Arnedt JT, Owens J, Crouch M, Stahl J, Carskadon MA. Neurobehavioral performance of residents after heavy night call vs after alcohol ingestion. Jama 2005;294:1025–33. https://doi.org/10.1001/jama.294.9.1025.
[125] Jokinen T, Saarenpaa-Heikkila O, Laippala P. A new visual performance test offers new variables for assessing daytime vigilance: short pauses in performance. Behav Res Methods Instrum Comput 1999;31:122–8.
[126] Kocher H, Warwick J, Al-Ghnaniem R, Patel A. Surgical dexterity after a "night out on the town. ANZ J Surg 2006;76:110–2. https://doi.org/10.1111/j.1445-2197.2006.03664.x.
[127] Hanks GW, O'Neill WM, Simpson P, Wesnes K. The cognitive and psychomotor effects of opioid analgesics. II. A randomized controlled trial of single doses of morphine, lorazepam and placebo in healthy subjects. Eur J Clin Pharmacol 1995;48:455–60.
[128] Kabat MH, Kane RL, Jefferson AL, DiPino RK. Construct validity of selected automated neuropsychological assessment metrics (ANAM) battery measures. Clin Neuropsychol 2001;15:498–507. https://doi.org/10.1076/clin.15.4.498.1882.
[129] Rebelo F, Soares MM. Advances in usability evaluation. CRC Press; 2012.
[130] Roth T, Richardson GR, Sullivan JP, Lee RM, Merlotti L, Roehrs T. Comparative effects of pravastatin and lovastatin on nighttime sleep and daytime performance. Clin Cardiol 1992;15:426–32.
[131] Starkey C. Athletic training and sports medicine: an integrated approach. Jones & Bartlett Publishers; 2012.
[132] Lowe C, Rabbitt P. Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: theoretical and practical issues. Cambridge neuropsychological test automated battery. International study of post-operative cognitive dysfunction. Neuropsychologia 1998;36:915–23.
[133] Lezak MD, Howieson DB, Bigler ED, Tranel D. Neuropsychological assessment. Oxford University Press; 2012.
[134] Woods SP, Scott JC, Dawson MS, Morgan EE, Carey CL, Heaton RK, et al. Construct validity of Hopkins Verbal Learning Test-Revised component process measures in an HIV-1 sample. Arch Clin Neuropsychol 2005;20:1061–71. https://doi.org/10.1016/j.acn.2005.06.007.
[135] Sánchez-Cubillo I, Periáñez JA, Adrover-Roig D, Rodríguez-Sánchez JM, Ríos-Lago M, Tirapu J, et al. Construct validity of the Trail Making Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. J Int Neuropsychol Soc 2009;15:438–50. https://doi.org/10.1017/S1355617709090626.
[136] Kelland DZ, Lewis RF. The Digit Vigilance Test: reliability, validity, and sensitivity to diazepam. Arch Clin Neuropsychol 1996;11:339–44.
[137] Jaeggi SM, Buschkuehl M, Jonides J, Perrig WJ. Improving fluid intelligence with training on working memory. Proc Natl Acad Sci USA 2008;105:6829–33. https://doi.org/10.1073/pnas.0801268105.
[138] Diehr MC, Cherner M, Wolfson TJ, Miller SW, Grant I, Heaton RK, et al. The 50 and 100-item short forms of the Paced auditory serial addition task (PASAT): demographically corrected norms and comparisons with the full PASAT in normal and clinical samples. J Clin Exp Neuropsychol 2003;25:571–85. https://doi.org/10.1076/jcen.25.4.571.13876.
[139] Woods SP, Delis DC, Scott JC, Kramer JH, Holdnack JA. The California Verbal Learning Test–second edition: test-retest reliability, practice effects, and reliable change indices for the standard and alternate forms. Arch Clin Neuropsychol 2006;21:413–20. https://doi.org/10.1016/j.acn.2006.06.002.
[140] Brown DT. Review of the Kaufman adolescent and adult intelligence test (KAIT). J. Sch. Psychol. 1994;32:85–99. https://doi.org/10.1016/0022-4405(94)90030-2.
[141] Block RA, Gruber RP. Time perception, attention, and memory: a selective review. Acta Psychol (Amst) 2014;149:129–33. https://doi.org/10.1016/j.actpsy.2013.11.003.
[142] Bilker WB, Hansen JA, Brensinger CM, Richard J, Gur RE, Gur RC. Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. Assessment 2012;19:354–69. https://doi.org/10.1177/1073191112446655.