TNO Kwaliteit van Leven

TNO-rapport

Arbeid
Polarisavenue 151
Postbus 718
2130 AS Hoofddorp

www.tno.nl/arbeid

T 023 554 93 93
F 023 554 93 94

Work hours fatigue and effects on patient safety. A review of simulator studies

KvL/031.20452/V&GW.2010.647/KAM.kos

Datum 25 december 2010

Auteurs Jakko van Kampen
Joyce Kramer
Jan Maarten Schraagen
Meike Smit
Robert Bezemer

Index

Summary

In healthcare continuous patient care is achieved by around the clock rostering of healthcare professionals (rotating shifts, on-call, extended shifts). Healthcare workers regularly need to work relatively demanding schedules whose characteristics may lead to fatigue and a reduced performance which could impact patient safety. In recent years there has been increasing criticism to work hours in medicine. IGZ has asked TNO Quality of Life to conduct a systematic review of existing literature on the potential effects of work hours on fatigue and patient safety. A broad and systematic literature review was conducted including research on working hours, fatigue and patient safety in the medical domain. The report builds upon four existing systematic reviews which were conducted by other authors. In addition it includes an in-depth look at a specific set of nine studies.

These nine included studies specifically examined the impact of fatigue on surgical performance as measured with surgical simulators. Overall there is convincing evidence that the (surgical) proficiency of surgeons in training is impaired under conditions of fatigue. This conclusion is in accordance with a considerable body of other research. Patient safety may be at risk when surgeons in training that are sleep-deprived and/or otherwise fatigued need to perform a laparoscopic surgery on a patient. However, there is insufficient data to conclusively prove this relationship or estimate its importance for patient safety in real life. More research would be needed to fully inform on the Dutch perspective. The simulator studies presented in this review suggest feasible methods that could be used to test the effects of different work schedules on surgical performance and the effectiveness of changes in work schedules for improving performance (without endangering patients).

1 Introduction

Patient care is needed not only during office hours but also at night and during the weekends. Around the clock availability of qualified personnel is therefore required in many healthcare professions. To achieve this availability a variety of rostering techniques is used. Common examples of this are rotating shifts, on-call shifts and extended shifts. Working within these rosters many healthcare workers such as nurses, residents and specialists, work a comparatively long number of hours which may be irregular and include night work. It is known from research in other domains that working in rosters with these characteristics may lead to fatigue and a reduced performance. Therefore, the question arises whether patient safety is compromised by these schedules.

Work hours in medicine have in recent years increasingly been subject to criticism. A central part of this critique is the potential impact of long work hours on fatigue amongst staff and its possible impact on patient safety. It is thought that the long work hours may lead to tired doctors making erroneous decisions and medical errors. These criticisms are heard in the Netherlands, in other European countries and in North America. In the Netherlands, a series of critical televised documentaries by Zembla was particularly influential on the general perspective of work hours and its effect on patient safety.

Legal limitations on work hours are the most common means to control the potential impact of work hours on (patient) safety. In European countries, the European Working Time Directive (EWTD) has limited the maximum number of work hours per week to 48 hours on average with the Netherlands and several other nations receiving an exception for working up to 52 hours per week until August 1st 2011[1;1]. This is a considerable reduction to historical hours of work, specifically for residents, as after a ruling by the European Court of Justice (2000) it includes the time spent whilst on-call. Whilst work hours have been restricted, it appears that in many cases compliance with these rules is limited. The Dutch labour inspectorate has started a 2010 inspection and compliance project on this issue [1] following earlier projects [2].

The relationship between work hours, fatigue, performance and human error has been extensively studied in many domains [3;4]. Most research, however, has focused on transportation or industrial sectors, for example amongst truck drivers in shipping [5] or in aviation [6]. It is not completely clear how these findings can be translated to the healthcare sector. While there is no reason to assume that healthcare workers are somehow physiologically 'different' than other workers, there could be differences in the effects of fatigue due to differing circumstances. It is clear that fatigue and associated errors are by no means solely a function of work hours. Many other factors can be important as well, for example: the time of day, regularity of work, intensity of tasks, breaks, home-work interference, and the characteristics of the task. Indeed, it is conceivable that a reduction in work hours might actually harm patient safety due to an increasing number of handovers and/or a reduction of time available for physician training.

1.1 Existing reviews

From the outset of the search it was known that there were already several published review articles related to this topic [7-10]. However, there were considerable differences in the focus and article selections used by these authors. Also, there were differences in the extent to which the reviews were systematic in that they clearly described the search process and the methodology for study inclusion and exclusion. Systematic reviews on the subject were published in 2004 [10] and 2008 [9]. In their reviews, Fletcher and colleagues concluded that there is insufficient evidence on patient safety to inform the process of reducing resident work hours. Ehara concluded that decrease of physician work hours is not harmful but favorable to patient safety. Notably, whilst both Fletcher and Ehara selected 7 studies for

use in their reviews and both had a similar period of investigation (1966 – 2004 versus 1966 – 2005) the overlap in their selections is limited to only one study. This illustrates that the search strategy and inclusion criteria used can strongly influence the results.

Recently, a further review was published [7]. Levine et al. examined the American Institute of Medicine's recommendations on eliminating workdays longer than 16 hours (extended work shifts). They retrieved 23 published and unpublished studies from the period of 1950 up to 2008 which examined the effects of reducing extended work shifts. Levine et al. concluded: *"... we found that reduction or elimination of resident work shifts exceeding 16 hours did not adversely affect resident education, and was associated with improvements in patient safety and resident quality of life in most studies"*.

In 2009 the American Institute Of Medicine's (IOM) 'committee on optimizing graduate medical trainee (resident) hours and work schedules to improve patient safety' published a book[8] which contains an extensive evaluation of the medical literature on work schedules in medicine. Some important conclusions made by the committee are that : *"There is ...extensive research that fatigue is an unsafe condition that contributes to reduced wellbeing for residents and increased errors and accidents."* and that there is *"considerable scientific evidence that 30 hours of continuous time awake ... can result in fatigue"*. The committee also concludes that the prevention of sleep deprivation is a fundamental way to optimize works schedules and to minimize fatigue. However there are *"simply too few data to reliably estimate the extent to which errors in performance by fatigued residents affects patients and cause them harm"*[8]. The committee notes that more research on the issue is needed but considers currently available evidence sufficient to recommend measures to reduce working hours in the U.S.

1.2 Research questions

Taken together the existing reviews conclude that excessive work hours can lead to fatigue amongst healthcare employees which may affect the occurrence of adverse events in patients. These reviews are based on a diverse set of studies. When studying the issue, measuring medical performance in real life is desirable because results are more easily generalized. However, experimental control is then difficult as potential confounding factors can make it difficult to draw firm conclusions. In addition, there are obvious ethical problems with exposing patients to potentially dangerous conditions.

Studies using validated medical simulators – most often surgical simulators - are a very interesting subset used to examine fatigue in medicine. Medical practitioners are partially or wholly deprived of sleep either by deliberate means (keep awake protocol) or naturally in the course of their work. Medical performance is then assessed in 'sleep deprived' and 'non sleep deprived' conditions. To measure performance, researchers have most often used laparoscopic virtual reality simulators but physical (full patient) simulators have been used as well. These studies have the advantage that they allow for comparatively good experimental control and sensitive, often automatically recorded performance measures can be used. In addition, because no actual patients are exposed to a sleep deprived doctor there are fewer ethical dilemmas. In light of the available literature and the different reviews that had already been conducted, it was decided to focus our review on this category of simulator studies. In total, the search process uncovered 9 published simulator studies. These studies were then included in a qualitative synthesis.

The Dutch Health Care Inspectorate (IGZ) has asked TNO Quality of Life to conduct a systematic review of existing literature on the potential effects of work hours on fatigue and patient safety. The main research question for this study is: *Does caregiver fatigue due to long and irregular work hours affect patient safety in the Netherlands?*

More specifically the research question for this review is: *Is there convincing evidence that long and/or irregular hours of work such as are common in Dutch medicine negatively impact performance on medical simulators?*

2 Method

A systematic review of scientific literature was performed to answer the research question. The scope of the review was limited to studies which were firstly conducted in a healthcare setting and secondly examined outcomes relevant for patient safety. In this review, we sought to gather evidence which either supports or refutes the relationships in our hypothetical model. The literature search, screening and review has been conducted with reference to the PRISMA framework [11;12].

2.1 Search model

To elucidate the aim of this study and to guide the search and selection process, we designed a hypothetical model for the search (Figure 1). Our goal was to retrieve relevant literature detailing studies which were conducted in a healthcare (hospital) setting and which included measurements and effect sizes for the hypothetical relationships in our search-model.

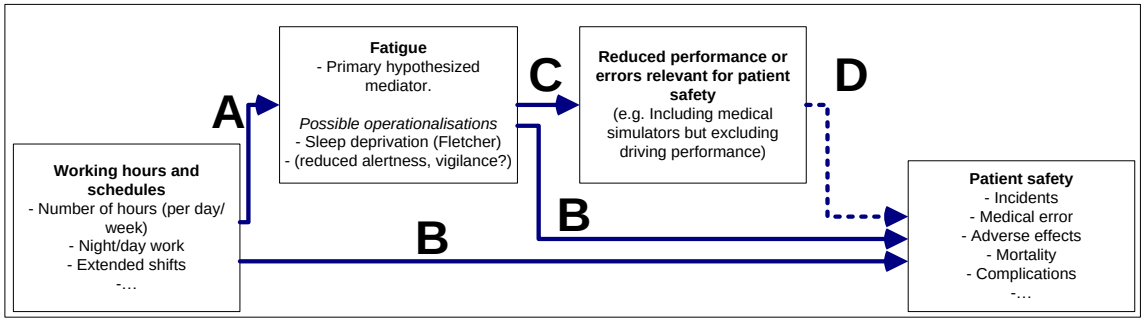

Figure 1, hypothetical model for search

The arrow labeled ‘A’ depicts studies examining the impact of work hours and schedules on caregiver fatigue. Studies labeled with ‘A’, however, did not measure the effect of caregiver fatigue on patient safety. The studies that did measure some aspect of patient safety, for example the frequency of incidents or the occurrence of medical error, are described with the arrows labeled ‘B’ These studies sometimes include direct measurements of caregiver fatigue, for example by using questionnaires. This may be examined as a mediating variable. Finally, the arrow labeled ‘C’ depicts studies that examine the impact of work hours and schedules on caregiver performance with a simulated task that closely resembles medical work. These tasks are relevant to patient safety but are not a direct measure. Experiments done with surgical simulators or realistic patient simulators are examples of this type of study. The validity of performance in these experiments for patient safety does need further study (dotted arrow D). The current search was aimed at the simulator studies depicted with arrow C.

2.2 Search strategy

For the literature search the SCOPUS database and search engine were used. The search was limited to the health sciences database (6800 journals) expanded with the psychology and human factors topic areas. The search terms were a cross-tabulation of keywords and synonyms describing the different aspects under examination, for example “working hours fatigue healthcare” (18 results) and “sleep deprivation patient safety” (143 results). A total of 21 search strings were used, a full listing of which is available in appendix A.

The 21 search strings yielded 1352 results. After the results were combined to remove duplicates 887 results remained. To this list were added a limited number of articles from additional sources, specifically the reference lists of other articles.

2.3 Selecting studies

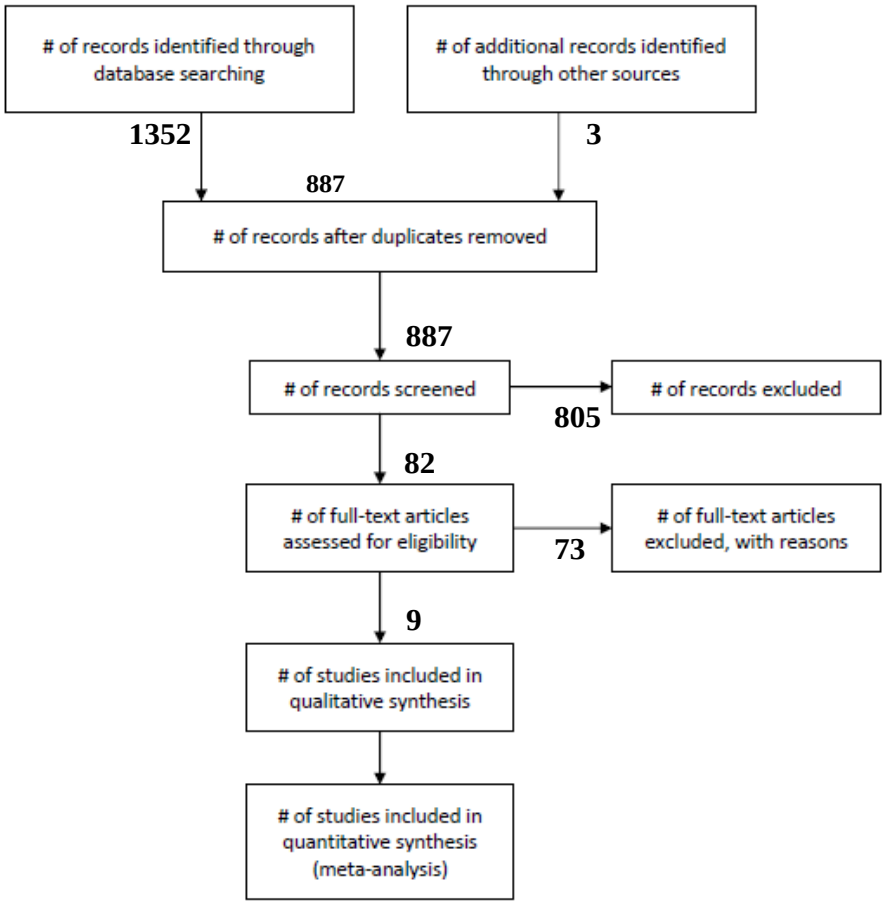

Figure 2, Summary of search results in PRISMA flowchart

The 887 selected abstracts proved to be a valuable collection. However, the collection needed to be extensively screened towards our research question. Articles were excluded when it could be determined from the abstract that the study design would not provide insight into the relationships of interest (Figure 1). To ensure consistency a log was kept of reasons for exclusion:

- Articles detailing fatigue as it occurs in patients (not in caregivers) and fatigue related to disease characteristics (for example research describing fatigue in patients suffering from cancer).
- Opinion pieces, editorials, notes and letters.
- Studies of fatigue and work hours conducted in other settings (shipping, industry, aviation).
- Studies of fatigue and work hours conducted in medical professions but not in hospitals.
- Studies in which outcome measures were not related to patient safety. For example outcomes detailing effects on injury to healthcare workers.
- Articles that examined seasonal differences in fatigue or learning and age effects (younger, older health care workers).
- Articles focusing on compliance to laws and regulation on work hours.

- Articles examining the effects of interventions aimed at preventing fatigue through other means than changes in work hours and scheduling (e.g., napping strategies).
- Articles examining the economic costs of alternative work hours, schedules and workloads.
- Development of (biomathematical) models and questionnaires to investigate fatigue.
- Articles which reported solely the perceptions of healthcare workers on the effect of fatigue on patient safety without including some form of measurement.
- Articles written in other languages than English.

Our focus was on individual studies. Systematic literature reviews and meta-analyses were not included as primary articles. A number of articles could not be adequately judged because there was no abstract available. These articles were fully retrieved for additional evaluation. After this step a list of 82 abstracts remained for full text retrieval and further study.

While studying full text articles studies were again excluded. These studies were primarily opinion pieces and articles on caregiver perceptions about work-hours and fatigue from which it was determined on closer reading that they did not contain actual measurements. This exclusion resulted in a list of approximately 40 studies which was further reduced to eight studies by the decision to focus exclusively on studies using medical simulators. This list of eight was then expanded with one study [13] through the studies reference lists to a total of nine studies.

3 Results

3.1 Overview of selected studies

In total nine articles were selected which described the effect of work hours on caregiver performance in a task that closely resembles medical work [4;13-20]. The tasks used included for example performance in a realistic patient simulator. All studies - except [14] and [13] - used shift timings to differentiate between high and low sleep deprivation and fatigue conditions. The most common setup was to conduct measurements in the morning both before a call period (pre-call) and after a call period (post-call). The types of shifts used in these cases are known as extended shifts: shifts that start in the morning and continue through a night of on-call duty with variable amounts of sleep. Leff et al[15] took a different approach: they examined (sequential) night shifts rather than extended shifts. Sharpe et al[14] and Taffinder et al[13] used more controlled circumstances to manipulate the amount of sleep in their participants. Sharpe kept participants awake over a 26-hour time span and conducted regular measurements within that period. Taffinder tested participants in the evening and in the morning and experimentally controlled the quality of the nights' sleep (undisturbed, disturbed and no sleep conditions).

Most studies were performed in a single hospital setting and were primarily focused on the performance of surgeons in training (surgical residents) and specialists training for other specializations (e.g. ICU residents). The number of participants included in the studies is relatively small (6-64 participants). Only Uchal[18] focused exclusively on surgeons (with nurses as a control group) and Gerdes[4] examined both surgeons and surgeons in training specifically. The nine studies show an overall increase in experimental realism over the years. Moving from relatively simple psychomotor tasks in rudimentary VR simulators (e.g. [13;19]) towards complex patient management tasks with a realistic patient simulator[14]. More recent research has used more complicated tasks which required the use of more diverse cognitive skills such as working memory, preparatory attention and 3-dimensional tracking ([4;16]). In Table 1 an overview of the included studies is presented with information on the study's method, the variables measured and the main results and conclusions.

Table 1 - Overview of evidence

Study	Country	Design	Subjects	Size	Simulator	Independent and mediating variables	Dependent variables	Performance decrement?
Sharpe, 2010	Canada	Experimental study (II-2)	ICU residents	N=12, one hospital	High fidelity patient simulator	Duration of wakefulness (26 hrs span, four measurements). Complexity of medical task.	Errors with patient safety implications.	Yes
Gerdes, 2008	USA	Prospective observational study (II-2)	Surgical residents and attending surgeons	N=14, one hospital	Laparoscopic simulator	Hours worked (pre-call vs post-call). Experience level.	Tool movement smoothness, hand movement smoothness, gesture proficiency, time to complete task, cognitive errors and psychomotor errors.	Yes
Kahol, 2008	USA	Prospective observational study (II-2)	Residents (multiple specialisations)	N=37, one hospital	Laparoscopic simulator.	Hours-worked (pre-call vs post-call).	Tool and hand movement smoothness, gesture proficiency, time to complete task, cognitive and psychomotor errors.	Yes

Study	Country	Design	Subjects	Size	Simulator	Independent and mediating variables	Dependent variables	Performance decrement?
Leff, 2008	United Kingdom	Prospective observational cohort study (II-2)	Surgical residents, residents of allied disciplines	N=21, one hospital	Laparoscopic simulator	Number of: sequential night shifts, steps taken, patients attended, calls received and self-reported sleepiness	Errors, time to complete task.	Yes
Jakubowicz, 2005	USA	Prospective observational study (II-2)	General surgical residents	N=8, one hospital	Laparoscopic simulator (ES3)	Hours worked (pre-call vs post-call), hours slept, caffeine, alcohol and cigarette intake	Errors, time to complete task, overall performance.	No
Uchal, 2005	Norway	Prospective observational study (II-2)	Surgeons and nurse controls	N=64, multiple hospitals	Physical simulator	Hours worked (post-work (8hr) vs post-call (24hr))	Accuracy error, tissue damage, leak rate, goal-non goal directed actions and time to complete task.	No
Eastridge, 2003	USA	Prospective observational study (II-2)	Surgical residents and trauma residents	N=35, one hospital	Laparoscopic simulator (MIST VR)	Hours worked (pre-call, on-call, post-call)	Errors, time to complete task.	Yes
Grantscharov, 2001	Denmark	Prospective observational study (II-2)	Surgical residents	N=14, one hospital	Laporoscopic simulator (MIST VR)	Hours worked (pre-call, post-call)	Time to complete task, error, unnecessary movement.	Yes
Taffinder, 1998	UK	Experimental study (II-2)	Surgeons in training.	N=6, one hospital	Laporoscopic simulator (MIST VR)	Morning and evening performance. Quality of nights sleep (undisturbed, disturbed, no sleep). Stress and arousal as mediators.	Total time, errors.	Yes

3.2 Evaluation of selected studies

Overall, of the nine studies examined, seven showed statistically significant effects of sleep deprivation on performance (Table 1). Two studies found results that seem to run counter to this view [17;18]. In Jakubowicz'[17] study the surgical residents in the sleep-deprived condition slept an average of 3.6 hours before the post-call trial which as the author notes is considerably longer than the sleep length attained in earlier studies. This together with the relatively small number of participants can explain that Jakubowicz did not find an effect[17].

The study by Uchal[18] was well-designed. The study involved a comparatively large number of participants (n=64), and included a separate control group of untrained nurses. The fact that no significant effect was found in Uchals' study is therefore noteworthy. There are, however, some limitations to the study. The study's post-call condition was not compared with a pre-call condition, as done by other authors[4;16;17;19;20], but was instead compared with a post-work condition. As noted by Leff[15] this is of relevance due to human circadian rhythms. The post-call measurements were done during a period of circadian upswing (8-9 a.m.), whilst the post-work measurements were done during a period of circadian low (4-5 p.m.). In addition, chronic sleep deprivation was not controlled in the study and may have influenced performance in the post-work condition negatively. These limitations could explain that no significant effect was found in this particular study.

An additional difference between Uchals' study and the other studies – with the exception of [4] - was that this study was conducted with experienced surgeons. Most often only surgeons in training were included in the studies. It is possible that the greater practical experience of the participants in Uchals' study allowed them to more adequately maintain performance under a condition of sleep deprivation. Only Gerdes included both surgeons and surgeons in training in a direct comparison[4]. Their results indicated that surgeons may be more able to maintain performance in a sleep deprived condition[4]. In Kahol's study more experienced residents showed better performance overall than less experienced residents but this difference disappeared when under sleep deprivation ([16]).

There has been a development in the study methodology over the years. Earlier studies tended to be mostly focused on psychomotor dexterity while the newer studies are increasingly showing the importance of both cognitive and psychomotor skills in tandem[4;14;16]. Sharpe[14] for example showed that while performance on a simple task showed improvement when tested over time, performance on more complex patient management tasks deteriorated. Kahol [16] showed that performance on cognitively demanding exercises suffers more strongly from sleep deprivation than performance on psychomotor tasks and also that psychomotor errors increase when conducting a parallel cognitive task. These studies suggest that whilst purely psychomotor tasks might be relatively unaffected by sleep deprivation this may not be the case with more realistic tasks which combine both cognitive and psychomotor components.

Table 2 includes a more in-depth evaluation of the individual studies. The table lists the studies' main conclusions, their strengths and limitations, and a view on their external validity.

Table 2 - Evaluation of studies

Study	Design	Notable limitations	External validity of outcome measures	Results and conclusions
Sharpe, 2010	Multiple performance measurements were made during 26 hours of continuous wakefulness under controlled standardized conditions. Blind evaluation of performance.	Scenarios were not yet fully validated, and residents were relaxing rather then working as they would be whilst on-call.	A high fidelity patient simulator was used. Various realistic tasks with high-cognitive complexity were undertaken. Previously defined outcomes that could have impact on patient safety were used. Outcomes were positively correlated with experience level. This study has the best apparent external validity of the simulator studies examined.	During prolonged continuous wakefulness of medical residents, clinical performance in the management of a simulated critically ill patient deteriorates. Complex scenarios were affected while less complex scenarios improved over time. Sharpe suggests that improvement in the less complex scenarios was likely a learning effect. The practice of scheduling residents for extended work shifts (>24 hrs) should be reconsidered.
Gerdes, 2008	Performance measures pre-call and post-call were taken. A comparison was made showing the effects of expertise (residents were compared with attending surgeons).	Exact number of hours worked and shift times not reported.	The tasks were more challenging then regular simulator tasks. Validation of the tasks was limited, but discriminative validity towards resident experience was shown in the pre-call condition.	Residents and attending surgeons showed significant decrement in proficiency post-call. Call-associated fatigue was associated with increased error rates in the cognitive skill domain, although less so in attending surgeons compared to their resident counterparts. Cognitive skills are more affected than psychomotor skills.
Kahol, 2008	Performance measures pre-call and post-call were taken. Analysis and tasks focused on both the cognitive and psychomotor skills needed for laparoscopic surgery.	Potential biases and limitations as well as the participant selection process are not clearly described. Further validation of outcome measures needed.	The tasks were conducted using a laparoscopic simulator but were more challenging then the regular simulator tasks used by older studies. Validation of the tasks was limited, but discriminative validity towards resident experience was shown in the pre-call condition.	Post call, in addition to increased cognitive errors, a significant decrease was recorded in the proficiency variables of memory, attention, and intermodal coordination tasks was shown. The time to complete tasks improved. There was a strong negative correlation between hours of sleep and cognitive errors. Effects were seen across the different resident skills levels examined.
Leff, 2008	The effect of 7 consecutive night shifts on performance was measured. Proficiency levels were equalized prior to the study to a predetermined benchmark level. Differences in work-intensity were measured. Within-subjects control one-week later during a week without night shifts.	It is possible that subjects were not adequately sleep deprived following first night shift. Poor performance may be due to other factors such as fears surrounding being tested.	Other aspects of clinical care (decision-making, cognitive abilities) were not investigated.	Newly acquired technical surgical skills deteriorate maximally after the first night shift, emphasizing the importance of adequate preparation for night rotas. Gradual improvement was observed across subsequent shifts towards baseline levels. This may be due to ongoing learning or adaptation to chronic fatigue.

Study	Design	Notable limitations	External validity of outcome measures	Results and conclusions
Jakubowicz, 2005	Pre-call post-call comparison using a within- subjects design. Training levels on simulator were equalized prior to study.	Post-call group had a relatively longer sleeping time (avg 3.6 hrs) compared to other studies (pre-call was 5.8). This has likely influenced effect size.	Jakubowicz considered the used ES3 simulator to be more realistic then MIST-VR. Cognitive deterioration of performance and effects on patient safety were not accounted for.	There was no diminution in performance before and after a 24-hour on-call period. There was a trend toward improved speed at the expense of accuracy. Jakubowicz notes a key distinction between this study and earlier studies with regard to the amount of sleep deprivation in the post-call group (3.6 hrs sleep on average).
Uchal, 2004	Between subjects design comparing two groups of surgeons. A control group of 64 nurses to validate outcome measures. Pre-tests were used to check for pre-experiment group-differences in skill level. The actual experiment was done with a physical simulator involving a foam stomach. 2 assessors blinded to time-of-day scored performance. Inter rater reliability was assessed.	Experimental power is reduced by comparing post-call with a post-work rather than pre-work condition. Testing at different times (circadian differences) could have counteracted effects of sleep deprivation. Chronic sleep deprivation and caffeine intake were not controlled.	Outcome measures have good apparent validity. 4 out of 6 outcome measures were shown to discriminate between nurses and surgeons.	Sleep deprivation had no impact on the studied outcome measures of a surgical task performed in a laparoscopic simulator.
Eastridge, 2003	Controls within subjects, three measurements were taken: pre-call, on-call and post-call. Learning effects were reduced by randomization of test sequence.	Cognitive skills, caffeine intake and chronic sleep deprivation are not accounted for.	Outcomes were validated performance measures in previous studies (MIST-VR).	Resident work schedules lead to sleep deprivation and fatigue. Call-associated sleep deprivation and fatigue are associated with increased technical errors and longer time to complete tasks in the performance of simulated laparoscopic surgical skills.
Grantcharov, 2001	Within subjects design with two measurements pre-call and on-call. Subjects had similar (limited) experience with the laparoscopic simulator.	Limited apparent control of potential confounds. Such as proficiency level, chronic sleep-deprivation, caffeine use, etc.	Performance on six tasks on the MIST-VR simulator (used in training). Tasks of progressive complexity.	Six different tasks were used. "Surgeons show impaired speed and accuracy in simulated laparoscopic performance after a night on call in a surgical department".
Taffinder, 1998	Pre-call and post-call performance was measured on six experimental nights. The quality of sleep nights was controlled in three conditions: undisturbed, simulated on-call (disturbed at three set times), and a no sleep condition. Participants were pre-trained on simulator.	Limited apparent control of potential confounds. Such as proficiency level, chronic sleep-deprivation, caffeine use, etc.	Laparoscopic performance, MIST-VR.	Significant decrement in performance across the conditions. Increased stress and decreased arousal across the conditions.

4 Discussion

Overall there is convincing evidence that the (surgical) proficiency of surgeons in training (residents) is impaired under conditions of fatigue. Seven of the nine studies that were included found a significant effect in this direction. The remaining two studies were found to have methodological limitations which could explain that no effect was found. This conclusion is in accordance with results from literature reviews in the medical domain[7-10] and in other domains[3] and with theoretical knowledge about performance, sleep and circadian rhythms.

These studies do not prove an actual impact of fatigue on patient safety, as we limited our review to simulator studies with no actual patients involved. However, the skills needed to perform these tasks were representative for surgical work. The researchers carefully chose to use representative tasks in their studies. This was done by using tasks that were previously validated for use in surgical training[19], by using validated psychological tests as a basis for tasks[4;16], or by demonstrating that more experienced operators were relatively better than less experienced users (discriminate validity).

There are, however, clear limitations to the included studies. The studies – with the exception of Leff[15] – are focused mostly on the effects of extended shifts, which means that they are less useful to evaluate the effects of hours of work that are less extreme. In the Netherlands and other European countries the European Working Time Directive has recently come into effect which should reduce the prevalence of extended shifts (though compliance appears to be difficult[1;2;8]). The simulator studies do however present feasible methods to test the effects of different work schedules and the effectiveness of changes in work schedules (without endangering patients).

Most of the studies used a relatively small sample size and selected their participants, surgeons in training, by convenience from a single hospital. The two studies that did include experienced surgeons suggest that they might be better able to cope with sleep deprivation than surgeons in training[4;18], although even experienced surgeons were not immune to the effects of sleep deprivation in Gerdes' study[4]. More research is needed to investigate the effects of experience on sleep deprivation.

An additional important limitation lies in the manipulation of fatigue through sleep deprivation. There is considerable variation in the amount of sleep obtained during on-call both within studies and across studies. Most studies relied on self-reported sleep length only, and none of the studies used more reliable methods such as actigraphy (automated recording of participant activity). Only Sharpe[14] and Taffinder[13] more fully controlled the amount of sleep in their participants. The benefit of measuring performance during normal shifts is that the levels of fatigue in these studies are likely to be comparable to fatigue experienced during actual work. However more controlled circumstances could do more to get information on safe limits for hours of work.

Better experimental control would also allow a researcher to isolate the effects of distinct factors such as time awake, time of day and work intensity. It is known that natural circadian rhythmicity strongly affects fatigue and performance across the 24-hours of the day with the strong circadian low point around 4 a.m. to 6 a.m. (the window of circadian low). Whilst most studies did choose to measure at roughly the same time each day, relatively wide windows were used and specifics were often not reported. None of the studies explicitly examined circadian effects. Finally, the work intensity of participants was not under experimental control in most studies. This may have affected the study results as the work intensity may vary considerably during day and night.

5 Conclusion

Overall there is convincing evidence that the (surgical) proficiency of surgeons in training is impaired under conditions of fatigue. This conclusion is in accordance with a considerable body of other research [3;7-10]. Patient safety may be at risk when surgeons or surgeons in training that are sleep-deprived and/or otherwise fatigued need to perform a laparoscopic surgery on a patient. However, there is insufficient data to conclusively prove this relationship or estimate its importance for patient safety in real life.

5.1 Dutch perspective

From the Dutch perspective it is important to note that the relevant research has for a large part been conducted in other nations, primarily the United States. These findings may thus not be fully applicable to the Dutch situation. Work practices - including work hours - tend to differ between Europe and the United States and also within Europe.

An important question is to what extent these findings are applicable to Dutch medical professionals. None of the studies investigated was performed in The Netherlands. It is clear from Table 1 that most studies focused on examining the effects of extended shifts. This type of scheduling appears to be less prevalent in the Netherlands. However the results still have relevance for the Dutch situation for several reasons. Compliance to the EWTD appears to still be far from guaranteed [1;2] and medical specialists such as surgeons are exempted from the EWTD.

Lastly compliance to the EWTD might not be sufficient guarantee for the prevention of fatigue. Studies that examined the effect of time awake found that after 18-20 hours awake performance was equivalent to an alcohol intoxication at the legal driving limit in the Netherlands[21;22]. Other studies found considerable impact of the chronic or cumulative fatigue that can arise from successive night shifts[23;24].

More research would be needed to fully inform on the Dutch perspective. The simulator studies presented in this review suggest feasible methods that could be used to test the effects of different work schedules and the effectiveness of changes in work schedules (without endangering patients).

Reference List

(1) Arbeidsinspectie. Projectplan Arbeids en Rusttijden 2010 Arts-assistenten. 2010.
Ref Type: Generic

(2) Arbeidsinspectie. Projectrapportage Arts-Assistenten 2007/2008. Project A845. 2008.
Ref Type: Generic

(3) Folkard S, Tucker P. Shift work, safety and productivity. Occupational Medicine 2003; 53(2):95-101.

(4) 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. American Journal of Surgery 2008; 196(6):813-820.

(5) Phillips R. Sleep, watchkeeping and accidents: a content analysis of incident at sea reports. Transportation Research Part F: Traffic Psychology and Behaviour 2000; 3(4):229-240.

(6) NTSB. Safety Study: A Review of Flight Crew Involved, Major Accidents of U.S. Air Carriers, 1978-1990. Washington, D.C: NTSB: 1994.

(7) Levine AC, Adusumilli J, Landrigan CP. Effects of reducing or eliminating resident work shifts over 16 hours: A systematic review. Sleep 2010; 33(8):1043-1053.

(8) Ulmer C, Wolmar DM, Johns MME. Resident duty hours. Enhancing sleep, supervision, and safety. The national academies press, Washington, D.C.; 2009.

(9) Ehara A. Are long physician working hours harmful to patient safety? Pediatrics International 2008; 50(2):175-178.

(10) Fletcher KE, Davis SQ, Underwood W, Mangrulkar RS, McMahon J, Saint S. Systematic review: Effects of residents work hours on patient safety. Annals of Internal Medicine 2004; 141(11):851-857.

(11) Moher D, Liberati A, Tetzlaff J, Altman DG, for the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009; 339(jul21_1):b2535.

(12) Liberati A, Altman DG, Tetzlaff J, Mulrow C, G++tzsche PC, Ioannidis JPA et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. PLoS Med 2009; 6(7):e1000100.

(13) Taffinder NJ, McManus IC, Gul Y, Russell RCG, Darzi A. Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. Lancet 1998; 352(9135):1191.

(14) Sharpe R, Koval V, Ronco JJ, Dodek P, Wong H, Shepherd J et al. The impact of prolonged continuous wakefulness on resident clinical performance in the intensive care unit: A patient simulator study. Critical Care Medicine 2010; 38(3):766-770.

(15) Leff DR, Aggarwal R, Rana M, Nakhjavani B, Purkayastha S, Khullar V et al. Laparoscopic skills suffer on the first shift of sequential night shifts: program directors beware and residents prepare. Annals of Surgery 2008; 247(3):530-539.

(16) Kahol K, Leyba MJ, Deka M, Deka V, Mayes S, Smith M et al. Effect of fatigue on psychomotor and cognitive skills. American Journal of Surgery 2008; 195(2):195-204.

(17) Jakubowicz DM, Price EM, Glassman HJ, Gallagher AJG, Mandava N, Ralph WP et al. Effects of a twenty-four hour call period on resident performance during simulated endoscopic sinus surgery in an Accreditation Council for Graduate Medical Education-compliant training program. Laryngoscope 2005; 115(1 I):143-146.

(18) 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. American Journal of Surgery 2005; 189(6):753-757.

(19) 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. American Journal of Surgery 2003; 186(2):169-174.

(20) Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Laparoscopic performance after one night on call in a surgical department: Prospective study. British Medical Journal 2001; 323(7323):1222-1223.

(21) Dawson D, Reid K. Fatigue, alcohol and performance impairment [5]. Nature 1997; 388(6639):235.

(22) Williamson AM, Feyer AM. Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. Occupational and Environmental Medicine 2000; 57(10):649-655.

(23) Dinges DF, Pack F, Williams K, Gillen KA, Powell JW, Ott GE et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep 1997; 20(4):267-277.

(24) Belenky G, Wesensten NJ, Thorne DR, Thomas ML, Sing HC, Redmond DP et al. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose-response study. Journal of Sleep Research 2003; 12(1):1-12.

Appendix A - Overview of search terms

	Keywords	Hits
1	Working hours safety fatigue	106
2	Working hours safety alertness	20
3	working hours patient safety	233
4	working hours adverse events	39
5	working hours medical error	140
6	working hours fatigue healthcare	18
7	working hours sleep healthcare	27
8	working hours alertness healthcare	5
9	Sleep deprivation patient safety	143
10	Sleep deprivation adverse events	24
11	Sleep deprivation medical error	151
12	Sleep deprivation safety fatigue	133
13	Sleep deprivation safety alertness	55
14	Working hours accidents healthcare	19
15	Fatigue management healthcare	144
16	sleep fatigue healthcare patient safety	14
17	sleep fatigue healthcare safety	28
18	sleep fatigue healthcare adverse events	3
19	sleep fatigue healthcare medical error	5
20	sleep fatigue healthcare incidents	3
21	sleep fatigue healthcare accidents	13
	Total	1352
	Total after removal of duplicates	887