Accident Analysis and Prevention 82 (2015) 20–26

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Managing fatigue: It really is about sleep

David Darwent *, Drew Dawson, Jessica L. Paterson, Gregory D. Roach, Sally A. Ferguson

Appleton Institute, Central Queensland University, 44 Greenhill Road, Adelaide, SA 5034, Australia

ARTICLE INFO

Article history:
Received 6 January 2015
Received in revised form 13 May 2015
Accepted 13 May 2015
Available online xxx

Keywords:
Sleep
Fatigue
Shiftwork
Biomathematical model
Fatigue risk management
Fatigue risk management system
Inter-individual differences

ABSTRACT

Biomathematical models of fatigue can assist organisations to estimate the fatigue consequences of a roster before operations commence. These estimates do not account for the diversity of sleep behaviours exhibited by employees. The purpose of this study was to develop sleep transfer functions describing the likely distributions of sleep around fatigue level estimates produced by a commercial biomathematical model of fatigue. Participants included 347 (18 females, 329 males) train drivers working commercial railway operations in Australia. They provided detailed information about their sleep behaviours using sleep diaries and wrist activity monitors. On average, drivers slept for 7.7 ($\pm$1.7) h in the 24 h before work and 15.1 ($\pm$2.5) h in the 48 h before work. The amount of sleep obtained by drivers before shifts differed only marginally across morning, afternoon and night shifts. Shifts were also classified into one of seven ranked categories using estimated fatigue level scores. Higher fatigue score categories were associated with significant reductions in the amount of sleep obtained before shifts, but there was substantial within-category variation. The study findings demonstrate that biomathematical models of fatigue have utility for designing round-the-clock rosters that provide sufficient sleep opportunities for the average employee. Robust variability in the amount of sleep obtained by drivers indicate that models are relatively poor tools for ensuring that all employees obtain sufficient sleep. These findings demonstrate the importance of developing approaches for managing the sleep behaviour of individual employees.

1. Introduction

In recent decades, the limitations of traditional hours of service regulations for the management of fatigue have been recognized (Jones et al., 2005; Sussman and Coplen, 2000). A consensus view is emerging that effective fatigue risk management systems (FRMS) may be achieved by policies and procedures implemented at the organizational and individual employee levels (Lerman et al., 2012). Proponents argue that a locally-tailored fatigue-risk management system can mitigate fatigue-related risks while concurrently allowing greater flexibility in the hours that employees can work (Cabon et al., 2012; Dawson and Zee, 2005; Gander et al., 2011).

In the FRMS framework, fatigue-related errors can be viewed as the end-point of a causal sequence of events termed, 'the fatigue hazard trajectory' (Dawson and McCulloch, 2005). The trajectory concludes with a fatigue-related error committed by an individual in a fatigued state, exhibiting symptoms or signs of fatigue. The fatigued state arises because insufficient sleep is obtained to maintain alertness at a given time of day after a given length of wake. Insufficient sleep is attributable to either an organisational failure to provide adequate rest opportunities or an individual failure to obtain sufficient sleep in an otherwise adequate rest opportunity.

The incidence of fatigue-related errors may be reduced via screening assessments targeted at sequential steps of the hazard trajectory (Dawson and McCulloch, 2005). Thus, biomathematical models of fatigue use software-based algorithms to assess whether scheduled work/rest periods provide employees with sufficient sleep opportunities (Mallis et al., 2004). The amount of sleep obtained by employees in these opportunities can be evaluated when they report for work, using either sleep monitoring devices or direct self-report. Symptoms and signs of fatigue that otherwise present at work can be detected using one or more fatigue recognition technologies (Balkin et al., 2011).

1.1. Biomathematical models of fatigue

Organisations implement biomathematical models to manage the fatigue-related risks associated with hours of work. The scores

* Corresponding author at: Appleton Institute, Central Queensland University, 44 Greenhill Road, Adelaide, SA 5034, Australia. Tel.: +61 8 8378 4515.
E-mail addresses: david.darwent@cqu.edu.au (D. Darwent), drew.dawson@cqu.edu.au (D. Dawson), jessica.paterson@cqu.edu.au (J.L. Paterson), greg.roach@cqu.edu.au (G.D. Roach), sally.ferguson@cqu.edu.au (S.A. Ferguson).

http://dx.doi.org/10.1016/j.aap.2015.05.009
0001-4575/

output by biomathematical fatigue models vary in their specific metrics, but all provide an estimate of the fatigue level associated with rostered shifts (Mallis et al., 2004). The assumption underlying numerical quantification is that higher levels of fatigue yield elevated levels of fatigue-related risk. Outputs are typically represented on a dimensionless numerical scale, i.e. without a unit of measurement, but some models include transfer functions that relate outputs to measurable phenomena, e.g. reaction time (Mallis et al., 2004). The purpose of these is to provide an external frame of reference for calibrating and interpreting model outputs.

The distinction between the fatigue level outputs produced by models and fatigue-related risk is important. On the one hand, 'fatigue level' refers to neurobehavioural deficits caused by disturbances in circadian and sleep homeostatic processes (Dijk and Archer, 2009). On the other, 'fatigue-related risk' refers to the extent of exposure to the potential costs of accidents caused by fatigue. The latter is the product of accident likelihood and the financial, social and human cost of accidents. In the absence of mitigating factors, fatigue level contributes to risk only by increasing the likelihood of fatigue-related accidents. By implication, the association between outputs and risk is unlikely to be constant across industries, nor within them.

Model outputs are typically used to classify shifts into tiered risk categories based on fatigue model outputs, e.g. safe vs. unsafe, low vs. moderate vs. high risk. Stratification of model outputs provides a more convenient method for precluding a given sequence of work or for imposing risk-mitigation strategies than raw scores alone. To account for variable risk profiles across industries, most commercial models permit organisations to modify the fatigue score thresholds for delimiting categories of risk. Despite this, industry reports and expert commentary continue to raise concerns in respect to the potential for over-reliance on fatigue models to evaluate safety risks (Civil Aviation Safety Authority, 2014; Dawson et al., 2011; Fourie et al., 2010; Gander et al., 2011; Independent Transport Safety Regulator, 2010).

One factor contributing to these concerns is the relative paucity of research to establish empirical thresholds for classifying shifts into risk categories (Williamson et al., 2011). To date, empirical research on fatigue models has focused primarily on validation of fatigue level estimates (Van Dongen, 2004). The presumption made by most models of a simple linear relationship between fatigue and the safety risks of work is not borne out by empirical investigation (Williamson et al., 2011). Establishing the nature of this relationship is problematic because of the low frequency of accidents in some industries, poor accident reporting and/or publication standards, and difficulties associated with causal attribution in accident investigation (Armstrong et al., 2013; Radun and Summala, 2004).

Another factor contributing to concerns with biomathematical models is that outputs are applicable only to the average individual. Fatigue model outputs are generated under the implicit assumption that the fatigue consequences of a roster are uniform for all employees. By implication, the amount of sleep obtained by employees in the rest periods of a given roster is presumed to be the same. In reality, some proportion of employees will obtain less sleep than predicted, with the consequence that fatigue score outputs are likely to underestimate the fatigue experienced by these employees. Conversely, another proportion will obtain more sleep than predicted, with the consequence that fatigue score outputs are likely to overestimate the fatigue experienced by these employees.

1.2. Purpose of this manuscript

The purpose of this investigation is to develop sleep transfer functions that permit interpretation of fatigue score outputs for a group of employees, i.e. rather than just the average employee. Like other transfer functions, sleep provides an intuitive metric for the non-expert to interpret fatigue level estimates. Thus, in the presence of doubt about the exact link between fatigue scores and fatigue-related risk, information about sleep could serve as useful supplementary information on which to base decisions. A sleep transfer function is potentially useful because it is one of the two basic factors that contribute to fatigue level estimates. Percentile distributions of sleep therefore provide a proxy for the likely distribution of fatigue level estimates for a group of employees working a given roster.

To quantify sleep, we use the prior sleep model proposed by Dawson and McCulloch (2005), which posits simple heuristics to evaluate whether an employee has obtained sufficient sleep before starting work. According to this model, employees should aim to obtain at least X h of sleep in the 24 h before work and Y h of sleep in the 48 h before work. The authors propose X and Y thresholds of 5 and 12 h, respectively, although subsequent empirical modelling of performance and error suggest that an X threshold of 6 h might better distinguish poor performance (Ferguson et al., 2011; Thomas and Ferguson, 2010). In this manuscript, the empirical relationship between the fatigue level outputs of a model and distributions of the sleep obtained by employees in the 24 and 48 h before work was evaluated. The proposed approach to extending the application of fatigue models is potentially generalizable to other models, but the Fatigue Audit InterDyne (FAID) software developed by our research group was utilised in this instance.

2. Material and methods

2.1. Ethics

The research protocol complied with the Australian National Guidelines on Ethical Conduct in Human Research. Ethical approval for conducting the studies was granted by the Human Research Ethics Committee of Central Queensland University.

2.2. Recruitment

Analyses were based on two data sets: (1) an original investigation conducted in 1995–97 (Roach et al., 2003); and (2) a repeat investigation conducted in 2010–12. The target population in both studies were train drivers working metropolitan and rural passenger and freight rail operations in Australia. Potential recruits in the original 1995–97 study were targeted at seven rail organisations at 1 of 14 depots located across five Australian states. Potential recruits in the 2010–2012 study were sampled from three rail organisations at 1 of 20 depots located across four Australian states. Participation was open to all train drivers employed by the collaborating organisations.

Recruitment sessions were arranged with collaborating rail organisations at local depots. Attendees were informed that the purpose of the investigation was to enhance biomathematical models of fatigue and that participation would involve measurement of sleep behaviour using activity monitors. Potential recruits were also informed that participation was voluntary, that any information collected would be de-identified and confidential, and that non-participation or withdrawal from the study would not influence future employment conditions. At the closure of each recruitment session, attendees were given an Information Sheet, Consent Form, General Demographic Questionnaire and a replied-paid envelope in which to return the signed Consent Form should they agree to participate in the study. Participants did not receive any financial incentive for completing the study.

2.3. Participants

The general population of Australian train and tram drivers comprises roughly 12,000 individuals (96% males, 4% females) (Long and Shah, 2013), with a mean age of 46 years (Logistics Training Council, Rail Transport, 2013). The combined sample in this study comprised of 347 (18 females, 329 males) train drivers, but information collected from 25 of these participants was not considered due to non-compliance with the study protocol. These data were excluded either because activity monitors were not worn or sleep/work diaries were not filled out. The remaining 322 drivers included 15 females (4.6%) and 309 males (95.4%). These participants had a mean age of 39.5 (±14.2) years and had been shift workers for 19.3 (±9.0) years.

2.4. Protocol

Drivers collected work and sleep information during the course of their normal employment duties. Work start and end times were recorded in a work diary. Reported times were cross-referenced with company records when inconsistencies or reporting errors were suspected. Bed-times and get-up times were likewise recorded in a sleep diary and then cross-validated using wrist activity data. Wrist activity monitors were worn on the non-dominant wrist. Participants were instructed to wear the activity monitor for the duration of the protocol and to complete dairy entries at the start and end of all work and sleep periods, respectively. This method of measuring sleep/wake behaviour has been extensively validated in laboratory and field settings (Ancoli-Israel et al., 2007).

In the 1995–97 study, participants completed the protocol for 2 weeks. Wrist activity levels were measured using Gaehwiler activity monitors (Gaehwiler Electronic, Hombrechtikon, Switzerland). The devices were configured to sample activity counts in 30-s epochs. Sleep diary records and activity records were cross-referenced to generate estimates of sleep timing and duration using Sleepwatch software (Mini-Mitter, Oregon, USA).

In the 2010–12 study, participants completed the protocol for 4 weeks. Wrist activity levels were measured using Philips-Respironics Actiwatch 2 activity monitors (Philips-Respironics, Oregon, USA). The devices were configured to sample activity counts in 1-min epochs. Sleep diary records and activity records were cross-referenced to generate estimates of sleep timing and duration using ActiWare software version 5.57 (Philips-Respironics, Oregon, USA).

2.5. Data analysis

Fatigue level estimates were generated using the Fatigue Audit InterDyne (FAID) software (InterDynamics, Adelaide, South Australia). The FAID algorithms (Dawson and Fletcher, 2001) are widely used in the Australian rail industry and have been validated against fatigue data collected in both laboratory and field-based studies (Roach et al., 2004a,b; Van Dongen, 2004). The FAID software generates fatigue level estimates for all shifts of a roster based on the work/rest history across the prior seven days. The range of possible fatigue level outputs extend from 0 to 243, but working time regulations in the Australian rail industry impose an upper limit of 90.

The workday was divided into three equal time periods. These included a morning period from 6:00 am to 1:59 pm, an afternoon period from 2:00 pm to 9:59 pm, and a night period from 10:00 pm to 5:59 am. Work periods were classified as morning, afternoon or night shifts based on the time period in which the majority of work hours were scheduled. Mean differences in 24-h and 48-h sleep history between shift types were tested using mixed-effects analysis of variance (ANOVA).

2.5.1. Sleep transfer functions

Shifts were classified into one of seven, ten-unit intervals based on fatigue level outputs, either: 20–29, 30–39, 40–49, 50–59, 60–69, 70–79 or 80–89. Sleep transfer functions were three basic statistical summaries of the 24-h and 48-h sleep history metrics. These included mean sleep history (±SD), percentile distributions of sleep history, and per cent violations of threshold values.

Mean sleep history across intervals were plotted and mixed-effects ANOVA were performed to test for differences. Percentiles (5th, 25th, 50th, 75th and 95th) for sleep history were calculated for each interval. A linear regression was then fitted through the percentile values of each interval. The fitted lines were plotted to produce percentile distribution charts of sleep across fatigue level intervals. Sleep history values were compared with X thresholds (either 5 or 6 h of sleep) and Y thresholds (either 12 or 14 h of sleep) of the prior sleep model. The percent frequency of shifts for which observed sleep amounts were below X h in the prior 24 h and Y h in the prior 48 h were tallied for each fatigue level category and then plotted.

The data sets used in the present analysis were not used in the development or parameterisation of the FAID algorithms.

3. Results

3.1. Shifts and work times

In total, participants worked 2728 driving shifts across the data collection periods, including 1214 morning shifts, 1104 afternoon shifts, and 410 night shifts. They worked an average of 41.2 (±10.9) h per week, made up by 16.9 (±10.9) morning shift hours, 15.3 (±9.9) afternoon shift hours, and 9.1 (±7.3) night hours.

3.2. Prior sleep/wake history

On average, participants slept for 7.7 (±1.7) h in the 24 h before work and 15.1 (±2.5) h in the 48 h before work. Table 1 reports the prior sleep amounts for morning, afternoon, and night shifts. Mixed-model ANOVA indicated significant differences across shift timing for 24-h, $F_{2,2704} = 6.66$, $p < .001$, and 48-h, $F_{2,2494} = 7.33$, $p < .001$, sleep history, respectively. However, the mean differences were marginal, equating to just 12 and 19 min of sleep in the 24 and 48 h prior to shifts.

3.3. Fatigue level intervals

Fig. 1 presents a frequency histogram for shifts classified into each fatigue level interval. The majority of shifts (95.5%) had fatigue level scores within the 20–89 range. Scores below this range accounted for 1.9% of shifts while scores above this range accounted for 2.6% of shifts (not shown).

Table 1
Sleep history prior to morning, afternoon and night shifts.

	Morning	Afternoon	Night
Sleep prior 24 h	7.5 (±1.7)	7.7 (±1.6)	7.7 (±2.1)
Sleep prior 48 h	15.0 (±2.5)	15.3 (±2.4)	15.1 (±2.9)

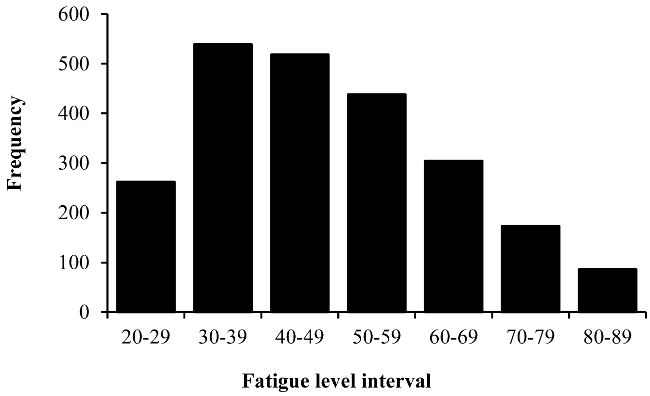

Fig. 1. Frequency histogram of shifts classified into 10-unit intervals of fatigue level scores.

3.4. Sleep transfer functions

3.4.1. Mean sleep history

Fig. 2 depicts the sleep history across fatigue level intervals. The 24-h sleep history (upper panel) ranged from 8.0 (±1.5) h in the lowest interval, i.e. 20–29, to 7.0 (±1.9) h in the highest interval, i.e. 80–89. The mean difference in sleep obtained equated to 64 min. The 48-h sleep history (lower panel) ranged from 15.8 (±2.4) to 14.2 (±2.8) h across the same categories, with a mean difference of 97 min of sleep. Mixed-model ANOVA indicated a significant main effect of interval for both metrics, $F_{7,2384} = 10.10$, $p < .001$ and $F_{7,2184} = 14.82$, $p < .001$, respectively.

3.4.2. Percentile distributions of sleep history

Fig. 3 illustrates fitted percentile distributions of sleep history across fatigue level intervals. The charts reveal a substantial range in the amount of sleep obtained by participants before shifts. The within-interval ranges, as indicated by the difference between the 5th and 95th percentiles, were substantially larger than the between-interval range, as indicated by the change in 50th percentiles between the lowest and highest intervals. The within-interval ranges also increased across interval categories as the absolute slopes ($|m|$) of the fitted percentile lines were not uniform. The fitted lines representing lower percentiles were successively steeper than those representing higher percentiles.

To illustrate the relationships described above by example, the difference in 24-h sleep history (upper panel) between the 5th and 95th percentiles in the lowest fatigue level interval (i.e. 20–29) was 5.0 h (the 90% CI ranged from 5.8 to 10.8 h.). In the highest interval

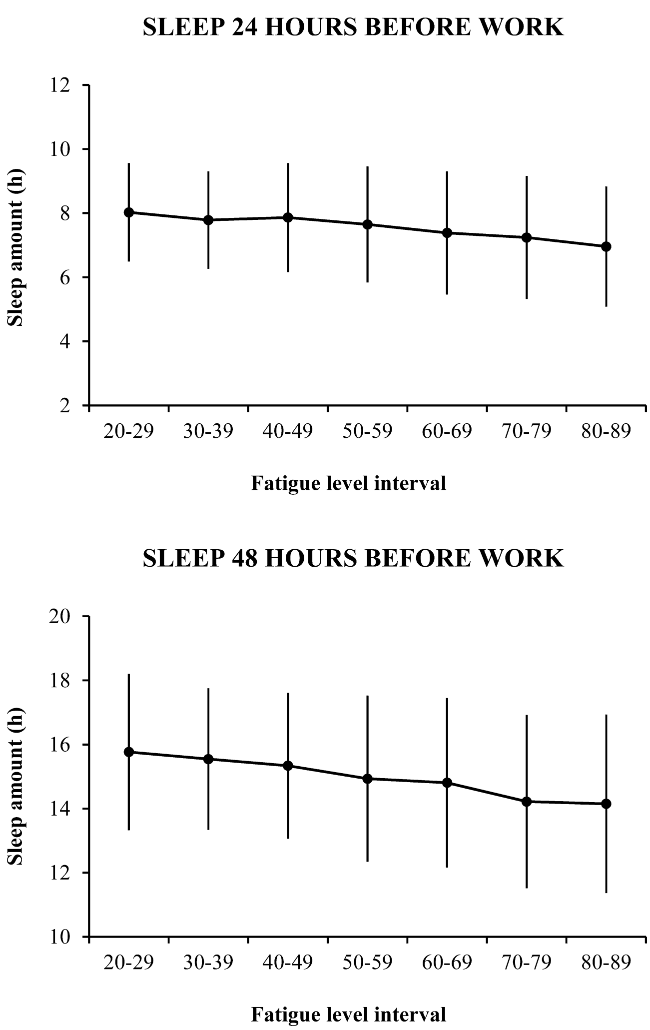

Fig. 2. Prior sleep/wake history for shifts classified into 10-unit intervals of fatigue level scores. The upper and lower panels show the mean amount of sleep obtained by drivers in the 24 and 48 h prior to work, respectively. Error bars denote standard deviations.

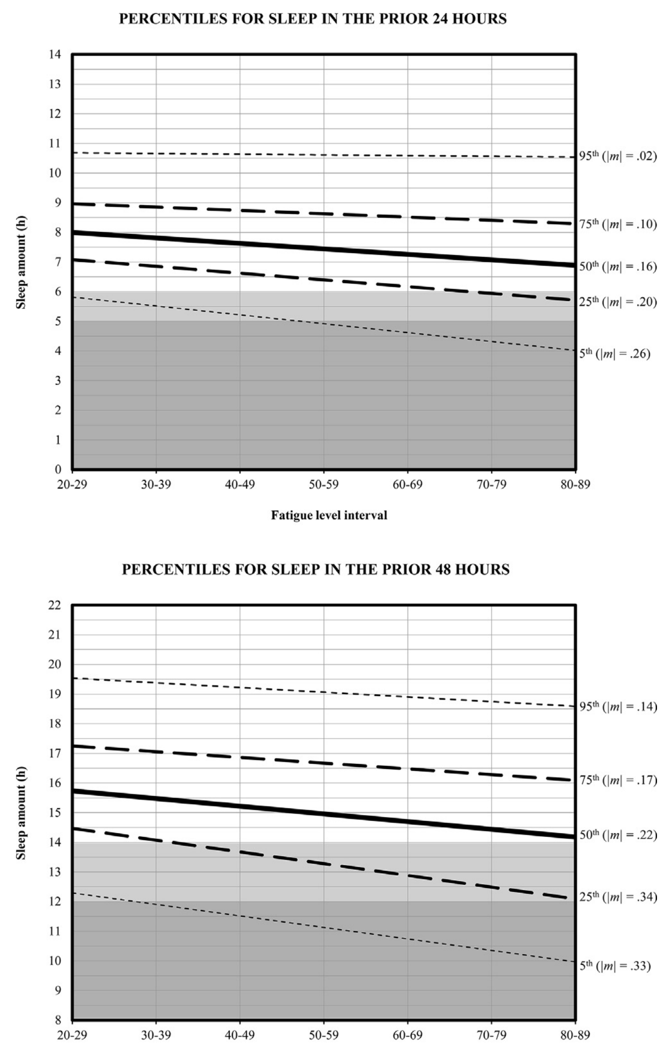

Fig. 3. Fitted percentile distributions of sleep in the 24 h (upper panel) and 48 h (lower panel) before work (5th, 25th, 50th, 75th and 95th percentiles). The absolute slope of each line, $|m|$, is indicated.

(i.e. 80–89), the difference was 6.3 h (the 90% CI ranged from 4.0 to 10.3 h). In comparison to these within-interval ranges, the between-interval range at the 50th percentile was only 1.3 h (ranging from 8.0 h in the lowest interval to 6.7 h in the highest interval). The between-interval range was not uniform across percentile levels. At the 95th percentile, the range was just 0.5 hours across intervals, whereas at the 5th percentile, the range was 1.7 h across intervals.

The difference in 48-h sleep history (lower panel) between the 5th and 95th percentiles in the lowest interval was also large at 7.4 h (ranging from 11.9 to 19.3 h, respectively). In the highest interval, the difference was 8.3 h (ranging from 10.1 to 18.4 h, respectively). At the 50th percentile, the between-interval range was 1.5 h (ranging from 15.8 h in the lowest interval to 14.3 h in the highest interval). Again, the between-interval range was not uniform across percentile levels. At the 95th and 5th percentiles, the between-interval range was 0.9 hours and 1.9 h respectively.

3.4.3. Violation of X *and* Y *sleep history thresholds*

Fig. 4 shows the percent of shifts where participants failed to obtain *X* hours of sleep in the 24 h. before shifts (upper panel) and *Y* h of sleep in the 48 h. before shifts (lower panel). A near-constant 2.5% of shifts violated an *X* threshold of 5 h when the fatigue level interval was below 50. Violations increased steadily across subsequent intervals to peak at nearly 15.0% of shifts. The percent of shifts violating an *X* threshold of 6 h was nearly double that across all intervals. Analogous distributions were obtained for *Y* threshold values, wherein ~5.0% of shifts had a 48-h sleep history below 12 h across fatigue level intervals below 50. The percent of violations increased across subsequent intervals and reached a maximum of nearly 25.0% of shifts in the 70–79 interval. The 14-h threshold yielded nearly double the number of violations across all intervals.

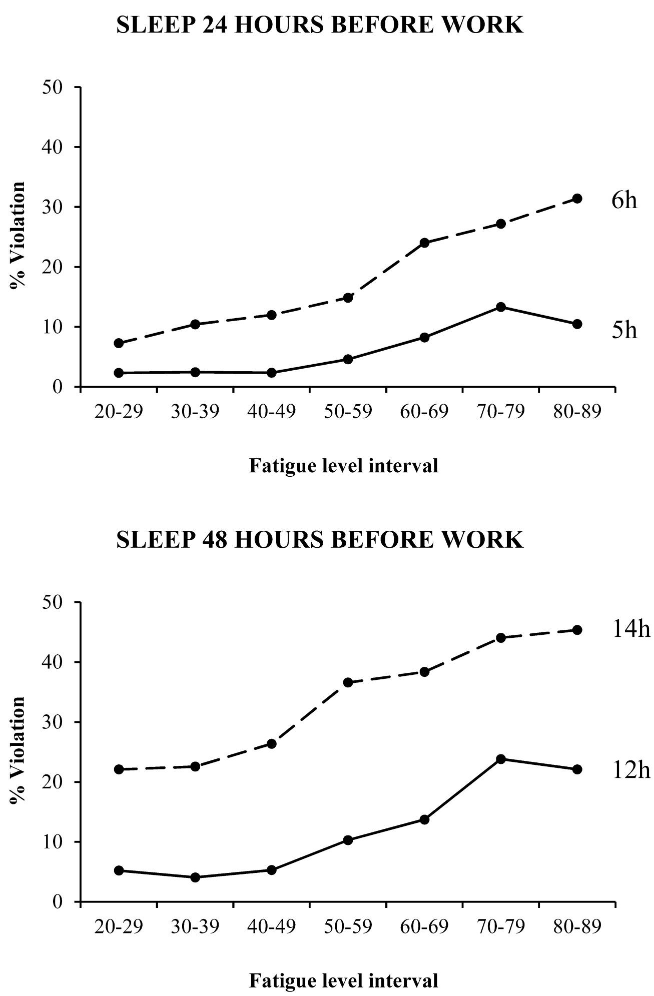

Fig. 4. Violation of prior sleep rules for shifts classified into 10-unit intervals of fatigue level scores. The top panel shows the per cent violation of the 24-h rule at two *X* threshold values (5 and 6 h). The lower panel shows the per cent violation of the 48-h rule at two *Y* threshold values (12 and 14 h).

4. Discussion

Software-based models of fatigue have typically been put forward as tools to evaluate the fatigue-related risks associated with workplace work/rest schedules. The original versions of these models lacked specific information enabling end-users to link fatigue level outputs with subsequent risk levels within a given industry. In the absence of empirical data, an alternative approach to mitigate potential safety issues was to benchmark fatigue output scores against standard roster patterns (e.g. standard 9 am–5 pm standard week, 4 × 12-h night shifts) or other patterns already accepted as safe by the industry (Dawson and Fletcher, 2001). Thus, a novel pattern of work could be considered safe if the calculated fatigue level scores were less than those for operations already considered safe. This historical legacy minimised radical change in the types of operations that were permitted under newly implemented fatigue-risk management approaches. However, it also obscured the distinction between fatigue level outputs, on the one hand, and the safety risks of operations, on the other.

Over-reliance on software-based models to evaluate fatigue-related safety risks is a commonly cited concern associated with fatigue-risk management systems (Civil Aviation Safety Authority, 2014; Fourie et al., 2010; Independent Transport Safety Regulator, 2010). This is because the link between fatigue levels and safety is not always linear, but is moderated by a host of mitigating variables including safety culture, duty tasks, and inter-individual differences across employees (Williamson et al., 2011). Fatigue may even have paradoxical effects on risk because employees sometimes exhibit risk mitigation strategies in recognition of an elevated potential for fatigue-related error (Gander et al., 2011). Community tolerance for risk may also vary across industry (Dawson and Zee, 2005). Thus, long work hours may be tolerated for some professions because the consequences of disrupted services outweigh the risk of fatigue-related accidents, e.g. medical doctors. The community is less likely to accept very long work hours for professions when the consequences of disrupted services are relatively benign, e.g. aviation pilots. For these reasons, setting fatigue score thresholds which distinguish between safe and unsafe work schedules is problematic for regulators. An alternative approach to risk classification is to instead focus on using models to manage the sleep behaviour of employees.

In this study, the relationship between fatigue model outputs and the amount of sleep obtained by employees in the 24 and 48 h before work was investigated. The sampled rosters were vetted for regulatory compliance using a software-based fatigue model before operations commenced. Unforseen events occasionally meant that drivers worked unplanned hours that were outside of regulatory limits, i.e. a fatigue level score >90. Otherwise, the assumption underlying compliance was that the average employee would have obtained sufficient sleep prior to shifts. This assumption proved valid in the current study, wherein drivers obtained an average of 7.7 h of sleep in the 24 h prior to shifts and 15.1 h of sleep in the 48 h prior to shifts. This is within the normal daily sleep range recommended for adults, i.e. 7.0–8.0 h per day (Ferrara and De Gennaro, 2001; National Institutes of Health, 2011). There was also a robust linear decrease in the amount of sleep obtained by employees across increasing fatigue level intervals. In comparison, only minor differences in prior sleep amounts were observed between morning, afternoon, and night shifts, despite well-established links between shiftwork operations and reduced sleep (Akerstedt, 2003; Kecklund and Akerstedt, 1995; Knauth et al., 1980). Taken together, these findings demonstrate the utility of fatigue models for designing round-the-clock rosters that provide sufficient sleep opportunities for the average employee.

There was substantial inter-shift variation in the amount of sleep obtained by employees prior to shifts. The variability was greater within fatigue level intervals than the mean differences across them by a factor of roughly five-fold. Trait-like inter-individual differences in physiological sleep parameters contribute to inter-individual variation in sleep behaviour (Tucker et al., 2007; Van Dongen et al., 2005). However, physiological contributions are not exclusive since shiftworkers exhibit intra-individual differences in sleep behaviour which are almost as great (Darwent et al., 2012; Dorrian et al., 2012). Thus, the variation in sleep amounts observed in this study most likely reflect the contribution of stable inter-individual differences in sleep physiology and idiosyncratic personal and social factors that influence sleep behaviour on a day-to-day basis. Given that sleep is one of the only two principle determinants of fatigue level (the other being circadian rhythmicity) (Borbely and Achermann, 1999), the observed differences in sleep amounts would be expected to yield corresponding differences in the fatigue levels experienced by employees. Hence, these findings suggest that within any given interval of estimated fatigue scores, the range of actual fatigue levels experienced by the employee group is likely to be large.

Fatigue-risk management is typically viewed as a shared responsibility (Dawson and McCulloch, 2005). Organizations are responsible for providing employees with sufficient sleep opportunities, while employees are responsible for getting adequate sleep in those opportunities. The study findings demonstrate that optimization of planned work/rest schedules using fatigue models has utility in regards to the former but is of limited value in ensuring the latter. Thus, pre-operational screening procedures could be improved if measures sensitive to inter-shift variability in employees' sleep behaviour were developed. One existing approach is to calculate fatigue scores for individual employees by inputting individual sleep times into a fatigue model (Reifman, 2004). The advantage of this approach is that both major physiological contributors to fatigue are taken into account. However, it implies a software interface that accepts novel sleep inputs and a system for capturing employees' sleep times. It is therefore best suited to small organisations with few employees, i.e. for ease of data entry, or large organisations with the capacity to automate data input and analysis e.g. Rangan et al. (2013). An alternative is to use the prior sleep history approach taken in this study—the advantage of which is the relative ease of implementation and the intuitive familiarity of the sleep metrics used.

According to the original prior sleep model, employees should obtain at least 5 h of sleep in the 24 h before work and 12 h of sleep in the 48 h before work (Dawson and McCulloch, 2005). While the association between reduced sleep time and elevated accident risk is well-established (Akerstedt, 2000; Drake et al., 2010; Philip and Akerstedt, 2006; Young et al., 1997), the notion that 'sufficient sleep' is a definable quantity is controversial (Van Dongen et al., 2003a). This is because the effect of sleep loss on fatigue is manifest in continuous dose–response deficits (Belenky et al., 2003; Van Dongen et al., 2003b), not at some physiological tipping point as the prior sleep model could be seen to imply. However, the objective in setting thresholds is not to define this tipping point, but rather to balance the safety-risks posed by sleep loss against unnecessary restriction on organisational efficiency and personal freedoms. Hence, a general set of sleep transfer functions, as presented in this study, would permit exploration of the organisational burden associated with different threshold values.

In this study, there were a proportion of shifts for which employees failed to meet the 5.0 and 12.0 h thresholds, irrespective of the sleep opportunities provided. The non-compliance rate rose above this baseline only for shifts with fatigue level scores beyond a score of 50. Below this, it was not possible to differentiate between reduced sleep caused by more demanding shiftwork schedules, on the one hand, and that caused by idiosyncratic day-to-day sleep disturbances, on the other. If these thresholds were accepted, then shifts with a fatigue level estimate above 50 should be targeted for further screening to identify employees with non-compliant sleep behaviours. Approximately 43% of planned shifts in this study fell into this range. Of these, roughly 15% were associated with non-compliance of one or both of these thresholds (i.e. roughly 6% of shifts in total). Higher sleep thresholds of 6.0 and 12.0 h were associated with a two-to-three fold greater rate of non-compliance across all interval ranges. If these thresholds were used instead, then shifts with a fatigue level estimate above 20 would be targeted for further screening because the frequency of non-compliance increased from this mark. Approximately 98% of planned shifts fell into this range, and of these, roughly 30% violated one or both thresholds.

The proposed approach attempts to ground the fatigue score predictions made by biomathematical models by plotting outputs against empirical distributions of shiftworker sleep. The extent to which the distributions of sleep observed in our population of train drivers can be generalized to other industries is unknown, but probably depends on the demographic profile of the industry under consideration. An analogous dataset could be used to generate sleep transfer functions for any given population or biomathematical model of fatigue. The usefulness of such functions would ultimately depend on the empirical relationship between predicted fatigue scores and observed sleep times. Hence, the applicability of this approach to novel populations in industries that use different biomathematical models requires further empirical investigation.

Knowledge of the prior sleep associated with mean fatigue scores would provide information on the likely distribution of fatigue scores around those means. One limitation in the application of this logic is the implicit assumption that individuals are equally susceptible to sleep loss. Previous research has demonstrated inter-individual differences in performance efficacy in response to equivalent levels of sleep debt (Van Dongen et al., 2004). In practice, this means that some individuals may exhibit fatigue-related performance deficits despite having obtained, in theory, an adequate amount of sleep. Conversely, others may exhibit normal performance despite having obtained a less than adequate amount of sleep. Identification of individuals who exhibit signs and symptoms of fatigue despite having obtained threshold levels of sleep remains fertile ground for ongoing research in fatigue safety management systems.

Biomathematical models of fatigue provide an estimate of the fatigue implications associated with a given work/rest schedule. Several recent reviews have identified limitations associated with fatigue prediction algorithms and the use of models in industrial settings (Gander et al., 2011; Dawson and Zee, 2005). In this manuscript, we propose that fatigue models should be viewed not as tools for quantifying fatigue-related risk, but rather as tools for mitigating risk through the identification of shifts where employees' sleep may be reduced. The advantage of an emphasis on empirical sleep distributions, as opposed to fatigue estimates based on work schedules, is that sleep information is readily appreciated by non-experts and safety evaluations would not rely solely on fatigue score predictions. Focusing on sleep prioritizes the importance of scheduling adequate sleep opportunities for all employees and for providing education, counselling and/or alternative scheduling arrangements where individuals demonstrate difficulty with compliance. Thus, in the absence of universally-applicable risk-based criteria for classifying model outputs, such as safe vs. unsafe, low vs. moderate vs. high, we suggest the more productive approach is to classify outputs on the basis of their implications for the distribution of employees' sleep, e.g. no aggregate sleep reduction vs. aggregate sleep reduction. We envisage that industry members would use this information to evaluate, based on industry-specific expertise and knowledge, whether a work schedule provides sufficient sleep opportunity for employees to perform work tasks safely.

Conflict of interest

Drew Dawson receives royalties from InterDynamics Pty Ltd. for a license associated with sleep opportunity prediction software (FAID^{TM}). No conflicts of interest are reported for David Darwent, Jessica Patterson, Greg Roach or Sally Ferguson.

Acknowledgements

Financial support for the study was provided by a CRC for Rail Innovation research grant—Next Generation Fatigue Models: Project number: R2.110.

References

Akerstedt, T., 2000. Consensus statement: fatigue and accidents in transport operations. J. Sleep Res. 9, 395.
Akerstedt, T., 2003. Shift work and disturbed sleep/wakefulness. Occup. Med. (Lond.) 53, 89–94.
Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., Pollak, C., 2007. The role of actigraphy in the study of sleep and circadian rhythms: American academy of sleep medicine review paper. Sleep 26, 342–392.
Armstrong, K., Filtness, A.J., Watling, C.N., Barraclough, P., Haworth, N., 2013. Efficacy of proxy definitions for identification of fatigue/sleep-related crashes: an Australian evaluation. Transp. Res. Part F: Traffic Psychol. Behav. 21, 242–252.
Balkin, T.J., Horrey, W.J., Graeber, R.C., Czeisler, C.A., Dinges, D.F., 2011. The challenges and opportunities of technological approaches to fatigue management. Accid. Anal. Prev. 43, 565–572.
Belenky, G., Wesensten, N.J., Thorne, D.R., Thomas, M.L., Sing, H.C., Redmond, D.P., Russo, M.B., Balkin, T.J., 2003. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose–response study. J. Sleep Res. 12, 1–12.
Borbely, A.A., Achermann, P., 1999. Sleep homeostasis and models of sleep regulation. J. Biol. Rhythms 14, 559–570.
Cabon, P., Deharvengt, S., Grau, J.Y., Maille, N., Berechet, I., Mollard, R., 2012. Research and guidelines for implementing fatigue risk management systems for the French regional airlines. Accid. Anal. Prev. 45, 41–44.
Civil Aviation Safety Authority, 2014. Biomathematical Fatigue Models. Civil Aviation Safety Authority.
Darwent, D., Dawson, D., Roach, G.D., 2012. A model of shiftworker sleep/wake behaviour. Accid. Anal. Prev. 45, 6–10.
Dawson, D., Fletcher, A., 2001. A quantitative model of work-related fatigue: background and definition. Ergonomics 44, 144–163.
Dawson, D., Ian Noy, Y., Harma, M., Akerstedt, T., Belenky, G., 2011. Modelling fatigue and the use of fatigue models in work settings. Accid. Anal. Prev. 43, 549–564.
Dawson, D., McCulloch, K., 2005. Managing fatigue: it's about sleep. Sleep Med. Rev. 9, 365–380.
Dawson, D., Zee, P., 2005. Work hours and reducing fatigue-related risk: good research vs good policy. JAMA 294, 1104–1106.
Dijk, D.J., LArcher, S.N., 2009. Light, sleep, and circadian rhythms: together again. PLoS Biol. 7, e1000145.
Dorrian, J., Darwent, D., Dawson, D., Roach, G.D., 2012. Predicting pilot's sleep during layovers using their own behaviour or data from colleagues: implications for biomathematical models. Accid. Anal. Prev. 45, 17–21.
Drake, C., Roehrs, T., Breslau, N., Johnson, E., Jefferson, C., Scofield, H., Roth, T., 2010. The 10-year risk of verified motor vehicle crashes in relation to physiologic sleepiness. Sleep 33, 745.
Ferguson, S.A., Paech, G.M., Dorrian, J., Roach, G.D., Jay, S.M., 2011. Performance on a simple response time task: is sleep or work more important for miners. Appl. Ergon. 42, 210–213.
Ferrara, M., De Gennaro, L., 2001. How much sleep do we need? Sleep Med. Rev. 5, 155–179.
Fourie, C., Holmes, A., Bourgeois-Bougrine, S., Hilditch, C., Jackson, P., 2010. Fatigue risk management systems: a review of the literature. Road Safety Research Report No. 10. Department of Transport, London.
Gander, P., Hartley, L., Powell, D., Cabon, P., Hitchcock, E., Mills, A., Popkin, S., 2011. Fatigue risk management: organizational factors at the regulatory and industry/company level. Accid. Anal. Prev. 43, 573–590.
Independent Transport Safety Regulator, 2010. Transport Safety Alert 34: Use of Biomathematical Models in Managing Risks of Human Fatigue in the Workplace. Independent Transport Safety Regulator.
Jones, C.B., Dorrian, J., Rajaratnam, S.M., Dawson, D., 2005. Working hours regulations and fatigue in transportation: a comparative analysis. Saf. Sci. 43, 225–252.
Kecklund, G., Akerstedt, T., 1995. Effects of timing of shifts on sleepiness and sleep duration. J. Sleep Res. 4, 47–50.
Knauth, P., Landau, K., Dröge, C., Schwitteck, M., Widynski, M., Rutenfranz, J., 1980. Duration of sleep depending on the type of shift work. Int. Arch. Occup. Environ. Health 46, 167–177.
Lerman, S.E., Eskin, E., Flower, D.J., George, E.C., Gerson, B., Hartenbaum, N., Hursh, S. R., 2012. Fatigue risk management in the workplace. Int. J. Occup. Environ. Med. 54, 231–258.
Logistics Training Council, Rail Transport, 2013. Logistics Industries Environmental Scan. Logistics Training Council for the Western Australian Department of Training and Workforce Development.
Long, M., Shah, C., 2013. Australian Transport and Logistics Industry: Forecasts of Labour and Skill Requirements 2013–17. Transport & Logistics Industry Skills Council.
Mallis, M.M., Mejdal, S., Nguyen, T.T., Dinges, D.F., 2004. Summary of the key features of seven biomathematical models of human fatigue and performance. Aviat. Space Environ. Med. 75, A4–A14.
National Institutes of Health, 2011. Your Guide to Healthy Sleep. US Department of Health and Human Services.
Philip, P., Akerstedt, T., 2006. Transport and industrial safety, how are they affected by sleepiness and sleep restriction? Sleep Med. Rev. 10, 347–356.
Radun, I., Summala, H., 2004. Sleep-related fatal vehicle accidents: characteristics of decisions made by multidisciplinary investigation teams. Sleep 27, 224–228.
Rangan, S., Bowman, J.L., Hauser, W.J., McDonald, W.W., Lewis, R.A., Van Dongen, H. P., 2013. Integrated fatigue modeling in crew rostering and operations. Can. Aeronaut. Space J. 59, 1–6.
Reifman, J., 2004. Alternative methods for modeling fatigue and performance. Aviat. Space Environ. Med. 75, A173–A180.
Roach, G.D., Fletcher, A., Dawson, D., 2004a. A model to predict work-related fatigue based on hours of work. Aviat. Space Environ. Med. 75, A61–A69.
Roach, G.D., Fletcher, A., Dawson, D., 2004b. Response to commentaries on a model to predict work-related fatigue based on hours of work. Aviat. Space Environ. Med. 75, A74.
Roach, G.D., Reid, K., Dawson, D., 2003. The amount of sleep obtained by locomotive engineers: effects of break duration and time of break onset. Occup. Environ. Med. 60, e17.
Sussman, D., Coplen, M., 2000. Fatigue and alertness in the United States railroad industry part I: the nature of the problem. Transp. Res. Part F: Traffic Psychol. Behav. 3, 211–220.
Thomas, M.J., Ferguson, S.A., 2010. Prior sleep, prior wake, and crew performance during normal flight operations. Aviat. Space Environ. Med. 81, 665–670.
Tucker, A.M., Dinges, D.F., Van Dongen, H.P., 2007. Trait interindividual differences in the sleep physiology of healthy young adults. J. Sleep Res. 16, 170–180.
Van Dongen, H.P., 2004. Comparison of mathematical model predictions to experimental data of fatigue and performance. Aviat. Space Environ. Med. 75, A15–A36.
Van Dongen, H.P., Baynard, M.D., Maislin, G., Dinges, D.F., 2004. Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. Sleep 27, 423–433.
Van Dongen, H.P., Maislin, G., Mullington, J.M., Dinges, D.F., 2003b. The cumulative cost of additional wakefulness: dose–response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 26, 117–129.
Van Dongen, H.P., Rogers, N.L., Dinges, D.F., 2003a. Sleep debt: theoretical and empirical issues. Sleep Biol. Rhythms 1, 5–13.
Van Dongen, H.P., Vitellaro, K.M., Dinges, D.F., 2005. Individual differences in adult human sleep and wakefulness: leitmotif for a research agenda. Sleep 28, 479–496.
Williamson, A., Lombardi, D.A., Folkard, S., Stutts, J., Courtney, T.K., Connor, J.L., 2011. The link between fatigue and safety. Accid. Anal. Prev. 43, 498–515.
Young, T., Blustein, J., Finn, L., Palta, M., 1997. Sleepiness, driving and accidents: sleep-disordered breathing and motor vehicle accidents in a population-based sample of employed adults. Sleep 20, 608–613.