Working Time Society consensus statements: Prescriptive rule sets and risk management-based approaches for the management of fatigue-related risk in working time arrangements

Kimberly A. HONN†, Hans P.A. VAN DONGEN†† and Drew DAWSON‡‡

1Sleep and Performance Research Center and Elson S. Floyd College of Medicine, Washington State University, USA
2Appleton Institute, CQUniversity, Australia

Received after WTS symposium comments: September 5, 2017
Received after editors’ revisions: January 19, 2018
Received after external review: June 21, 2018
Received after expert panel (Final accepted): January 3, 2019
Published online in J-STAGE January 31, 2019

Abstract: Traditionally, working time arrangements to limit fatigue-related risk have taken a prescriptive approach, which sets maximum shift durations in order to prevent excessive buildup of fatigue (and the associated increased risk) within shifts and sets minimum break durations to allow adequate time for rest and recovery within and/or between shifts. Prescriptive rule sets can be successful when, from a fatigue-related risk standpoint, they classify safe work hours as permitted and unsafe work hours as not permitted. However, prescriptive rule sets ignore important aspects of the biological factors (such as the interaction between circadian and homeostatic processes) that drive fatigue, which are critical modulators of the relationship between work hours and fatigue-related risk. As such, in around-the-clock operations when people must work outside of normal daytime hours, the relationship between regulatory compliance and safety tends to break down, and thus these rule sets become less effective. To address this issue, risk management-based approaches have been designed to regulate the procedures associated with managing fatigue-related risk. These risk management-based approaches are suitable for nighttime operations and a variety of other non-standard work schedules, and can be tailored to the particular job or industry. Although the purpose of these fatigue risk management approaches is to curb fatigue risk, fatigue risk cannot be measured directly. Thus, the goal is not on regulating fatigue risk per se, but rather to put in place procedures that serve to address fatigue before, during, and after potential fatigue-related incidents. Examples include predictive mathematical modeling of fatigue for work scheduling, proactive fatigue monitoring in the workplace, and reactive post-incident follow-up. With different risks and different needs across industries, there is no “one size fits all” approach to managing fatigue-related risk. However, hybrid strategies combining prescriptive rule sets and risk management-based approaches can create the flexibility necessary to reduce fatigue-related risk based on the specific needs of different work environments while maintaining appropriate regulatory oversight.

Key words: Fatigue risk management, Workplace, Alertness, Safety, Duty hours, Hours of service regulations

†Shared senior authorship.
*To whom correspondence should be addressed.
E-mail: kimberly.honn@wsu.edu

©2019 National Institute of Occupational Safety and Health

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (by-nc-nd) License. (CC-BY-NC-ND 4.0: https://creativecommons.org/licenses/by-nc-nd/4.0/)
Consensus Statements

1) Historically, prescriptive rule sets specifying maximum shift durations, minimum break durations and aggregated total work hours were developed in order to limit health risks due to physical fatigue. While well-entrenched as a regulatory approach, prescriptive rule sets are not well suited to managing the risks associated with mental fatigue.

2) Ideally, approaches to regulating the fatigue-related risk associated with working time arrangements should ensure regulatory compliance improves safety outcomes.

3) Prescriptive rule sets tend to be least effective when working time arrangements involve work during the typical range of nocturnal sleep (i.e., between approximately 9pm and 9am). This is due to a variety of factors, including:
   (a) the critical role of circadian rhythms in regulating sleepiness/alertness;
   (b) the cumulative effects of sleep restriction typically associated with daytime sleep on mental fatigue.

4) Risk management-based approaches (e.g., Fatigue Risk Management Systems; FRMS) have considerable potential to be more effective and more flexible than prescriptive rule sets for mitigating the risks associated with mental fatigue.

5) An FRMS (as part of a broader Safety Management System) is typically comprised of:
   (a) a policy and governance framework;
   (b) a training and education program;
   (c) an evidence-based risk assessment and mitigation methodology;
   (d) a monitoring and review process to ensure continuous improvement in compliance and efficacy.

6) FRMS governance of working time arrangements should involve a “shared responsibility” framework for managing fatigue-related risk at work in accordance with “just culture” principles. In broad terms, the employer should be primarily responsible for managing work-related causes of fatigue and the employees for managing non-work-related causes of fatigue. Where an employee (or supervisor) believes a worker may be unable to work safely, there should be a documented process whereby they are able to (and required to) notify the organization, determine the cause, and mitigate the risk.

7) Monitoring and review of an FRMS should involve:
   (a) quantifying the organizational risk and safety outcomes of fatigue including leading indicators and/or ‘near miss’ events;
   (b) developing and reporting valid and reliable key performance indicators;
   (c) assessing compliance with FRMS procedures;
   (d) assessing the effectiveness of, and subsequently updating of the FRMS procedures.

8) In determining how best to manage fatigue-related risk, an organization needs to consider:
   (a) the broader safety culture of the organization and societal context;
   (b) the resources, infrastructure and expertise available for implementing the FRMS;
   (c) the views of relevant stakeholders (e.g., employees and their representatives, relevant regulatory agencies, local communities, etc.);
   (d) the likely cost/benefit ratio of a proposed FRMS.

9) Over time, hybrid frameworks may evolve that permit organizations to utilize the benefits of both prescriptive and risk-based approaches in their FRMS.

10) The optimal fatigue risk management approach, whether it be prescriptive, risk management-based, or a hybrid approach, depends upon the sophistication and maturity of the given organization’s broader safety system/culture.

Consensus statements review expert panel: Stephen POPKIN¹ (Chair), Julie BULLAS², Steven R. HURSH³

¹Department of Transportation Volpe Center, USA
²Office of the National Rail Safety Regulator, Australia
³Institutes for Behavior Resources, USA

Full consensus among panel members on all statements.
Introduction

This manuscript is part of a series of consensus papers developed by the Working Time Society, as commissioned by the International Commission on Occupational Health. The goal of this series is to provide guidance for a broad, international audience of researchers, industry representatives, workers, labor representatives, policy makers, and other stakeholders on managing fatigue associated with non-standard working hours and ensuring worker health and safety. Collectively, the papers provide overviews of the current state of research, identify health and safety risks, make recommendations for effective interventions, and suggest future research directions. Each paper presents a number of consensus statements, developed through the procedures outlined in Wong et al.\(^1\), and describes the background information on which the consensus statements are based. The present paper is concerned with the management of fatigue-related risk in shift work systems and, more broadly, in working time arrangements (WTAs).

Shift work systems are essential for operating a wide range of around-the-clock industries, including manufacturing, resource extraction, medical care, law enforcement, and transportation. Many shift workers face circadian misalignment, that is, having to sleep during the daytime when the circadian rhythm of pressure for wakefulness is high and having to work during the nighttime when the circadian rhythm of pressure for wakefulness is low. Others face sleep restriction from extended or irregular shifts. Shift workers are therefore susceptible to sleep loss and fatigue-related impairment\(^2\).

Fatigue increases the potential for errors\(^3\). Fatigue-related errors may be inconsequential, particularly when there are automated alerts or co-worker cross-checks in place. However, a fatigue-related error aligned with certain situational factors—e.g., unexpected circumstances, system break-downs, additional hazards, or failing automated alerts—can have disastrous results\(^4\). Thus, an increase in errors due to fatigue produces a higher potential for incidents and accidents\(^5\).

In most operational settings, to limit the likelihood of fatigue-related errors, WTAs are constrained to help prevent excessive build-up of fatigue. Historically, fatigue has been managed primarily using prescriptive rule sets that limit WTAs based on specified maxima and minima for shift duration and break length. In recent years, WTAs that seek to assess and manage the risk of fatigue-related errors more directly have gained popularity. Here we discuss the context and review the advantages and disadvantages of both approaches.

Working Time Arrangements Based on Prescriptive Rule Sets

The primary means by which prescriptive rule sets seek to reduce the likelihood of fatigue-related errors is to specify maxima for shift duration and minima for time off between shifts. There may also be prescriptive limits on other dimensions of the WTAs; these are more variable across industries but can include paid rest breaks, minimum duration of rest breaks, maximum total weekly and monthly hours, maximum consecutive work days and/or night shifts, minimum days off between shift rotations, overtime restrictions, etc. Such rule sets have typically been developed over many decades, based on the following:

- Labor contract agreements between employers and employee representatives (e.g., unions), to varying degrees based on or augmented by information from personal experience. Such agreements are sometimes supplemented by expert scientific opinion derived from the (often limited) research literature.
- Regulatory control whereby an agency is charged with setting out the prescriptive rule sets that govern the WTAs across a large group of firms. In this case, all organizations under the jurisdiction of the regulator are required to comply with the same prescriptive rule set. These arrangements are typically negotiated between a regulator and an industry association.
- Mixed modes, where labor contracts typically further constrain the WTAs within the broader rules globally issued by the regulator. This can either serve to control fatigue even further, or can serve psychosocial goals that may or may not be related to fatigue. Thus, firms may have tighter constraints on working times imposed on them by their labor contract than by the regulator.

Initially, regulatory reforms in the 19th century were focused on the control of physical rather than mental fatigue. Because physical fatigue accumulates and discharges as a roughly linear function of time on and off task, respectively, WTAs were typically constrained using shift maxima and break minima. Shift maxima served to constrain the accumulation of fatigue beyond a certain level, and break minima were intended to provide sufficient rest to permit recuperation. The use of intra-shift breaks, shift duration maxima, and break duration minima provided simple yet appropriate ways to mitigate the risks associated with physical fatigue.

Industrial Health 2019, 57, 264–280
Accordingly, from the late 19th century until the 1930s, many industrialized countries developed regulatory frameworks and labor contracts that constrained (primarily physical) fatigue using simple rule sets based on multiple dimensions of “time on” and “time off” task. By the time of World War II, regulators, unions, and managers had been immersed in a 50 yr discourse about how practices may reduce fatigue risk using prescriptive rule sets. This was typically based on intra-shift breaks, shift maxima, and break minima, along with aggregate maximum hours over weekly to monthly periods.

During and following World War II, the concept of mental fatigue or sleepiness became a clear focus of military and industrial safety specialists and, more gradually, a focus for university-based research efforts. The growing reliance on night shift work and the increasing cognitive complexity of many jobs made employees even more susceptible to errors due to mental fatigue (or sleepiness). Mental fatigue increasingly became a focus of workplace safety discussions. For many organizations, the emerging risks, workplace errors, and accidents indicated an additional reason to constrain the WTAs to reduce the likelihood of mental fatigue-related errors.

Despite important differences in the build-up of fatigue risk associated with physical versus mental fatigue, stakeholders (including labor, management, and regulatory actors) often defaulted to the use of pre-existing regulatory levers (developed for physical fatigue) to address the issue of mental fatigue. The reasons for this are complex, but in many cases stakeholders preferred simple rules specifying shift maxima and break minima because they were familiar and compliance was easily assessed. Despite obvious limitations (discussed below), the use of prescriptive rule sets has persisted, primarily due to custom and practice rather than efficacy or evidence. As H.L. Mencken famously observed, “There is always a well-known solution to every human problem—neat, plausible, and wrong.”

**Features of Prescriptive Rule Sets**

Prescriptive rule sets are a common element of the regulatory landscape for fatigue in most developed economies. The scientific evidence base for prescriptive WTAs varies widely across industries—see references for reviews and guidance materials. In this paper, we have resisted the temptation merely to list the various prescriptive systems for managing fatigue-related risk. Not least because there are so many of them, but also because they are often variations of a common text, with each firm and regulator drawing upon pre-existing regulatory texts and (re) modifying them to meet local political, industrial, and regulatory circumstances at the time of their development.

It can be instructive to look at the bibliographies of these rule sets. They tend to feature as many, if not more, references to other regulations than to the scientific literature. Selective interpretation and over-generalization occur frequently. Sometimes, rules that appear ubiquitously in the regulatory texts may have only one or two studies with very limited generalizability underpinning them (for an illustrative discussion of these biases with respect to 12-h versus 8-h shifts, see Ferguson and Dawson).

As such, the evidence-base for prescriptive rule sets is typically limited. In many cases, there is either insufficient evidence or the existing evidence is too confounded to draw independent conclusions about individual features. Thus, averages, ranges, and/or consensual values for prescriptive rules (such as maximum shift duration) are relatively uninformative (especially when taken out of context), and we therefore do not list them here. Instead, we have chosen to identify the most common features of prescriptive rule sets, along with recent trends in prescriptive WTAs.

As indicated above, the primary features of prescriptive rule sets for WTAs typically include:

- Maximum shift durations
- Minimum break durations during shifts
- Minimum break durations between shifts
- Minimum days off per week/month/quarter/year
- Maximum total hours per day/week/fortnight/month/year
- Maximum consecutive working days or shifts

Secondary features may include:

- Maxima for time on task within a shift
- Minima for time off task (i.e., breaks) within shifts
- Minima for time off work between shift sequences (duty cycles)
- Differentiation of features based on sleeping location (e.g., residential community, motel, camp, on-site, in-vehicle)
- Differentiation of shift durations based on task load
- Differentiated shift durations for primary tasks (e.g., driving, flying) versus other duties
- Differentiation of shift and break durations based on time of day
- Differentiated aggregate total hours per week/fortnight/month based on time of day
- Differentiated minima for time off based on time of day
• Constraints on duration and frequency of overtime

Additional features that sometimes form part of prescriptive rule sets include:

• Requirement for relevant staff to undertake appropriate training in identification and mitigation of fatigue-related risk
• Increased employee responsibility for appropriate use of breaks to ensure adequate recovery (i.e., limits around secondary employment and leisure activities that reduce the recovery value of breaks)
• Incorporation of extended commuting times as an employer responsibility
• Differentiated task assignments based on time of shift (or during overtime) and/or based on time of day
• Provisions for providing on-site rest or sleep opportunities or arranging transportation for an employee deemed to have exceeded a safe work duration
• Self or peer assessment of fatigue prior to or during work and the requirement for an employee to notify the employer if they have reasonable grounds to believe they may be unfit for duty

In part because many of the prescriptive rule sets around WTAs have been developed based on principles other than just (physical) fatigue, there is a tendency for the rule sets to be overreaching and incorporate a range of other hazards indirectly linked to fatigue. These can be important influences on fatigue and safety (e.g., sleep disorders) but, in many cases, might be more appropriately dealt with as separate hazards rather than indirectly through the working time rules. Features of this nature that may be part of prescriptive rule sets (but might well be inappropriately placed there) include:

• Leave requirements
• Hydration and diet
• Heat stress
• Physical activity
• Noise and vibration
• Hazardous chemical exposure
• Medication use
• Sleep disorders

While it is important that prescriptive rule sets for WTAs cross-reference relevant indirect sources of risk, it may be more effective to manage them through other rules, standards or processes within the broader safety policy framework so that they can be addressed directly. For example, while dehydration can affect fatigue, it may be better addressed directly through a hydration policy with cross-referencing in the WTAs policy rather than indirectly through the working time rules. Similarly, medication use and sleep disorders may be better addressed through a medical fitness standard with cross-referencing in the WTAs policy rather than through the WTAs policy alone.

While the primary focus of prescriptive rule sets is to restrict the build-up of fatigue during and across shifts, in recent years, there appears to have been a trend toward emphasizing adequate recovery time\(^{[19]}\). For example, in the Australian National Heavy Vehicle Regulator (NHVR) Advanced Fatigue Management framework\(^{[20]}\), the risk associated with WTAs is determined based on the capacity to recover rather than the accumulation of fatigue per se. From this perspective, the WTAs are based on the provision of adequate time for recovery, rather than only limiting working time.

**Incongruence between Compliance and Safety**

The initial momentum for prescriptive rule sets emerged in the late 19th century as a result of broader labor reforms and reflected a complex (and often contradictory) combination of social, financial, and safety influences on negotiations between labor, management, and regulators—sometimes, but not always, informed by research evidence. At the time, six 12-h shifts per week were common in many manufacturing, mining, and service industries. Over the next 100 yr, the general trend across the developed world was toward sequences of four to six 8-h shifts\(^{[21]}\). By the late 1960s, most workers in developing countries had conformed to an 8-h work day and 40-h work week. Consequently, most shift systems divided the day into three 8-h shifts (morning, afternoon, night), with employees working 4 to 6 shifts sequentially followed by 2 to 3 d off.

From about 1980 onward, “reforms” across much of the developed world resulted in a reversal of this trend and a significant intensification of WTAs—especially in English-speaking countries. Many industries moved toward longer shifts, longer hours, and greater work intensification\(^{[22, 23]}\). This trend has been associated with growing concern over an increased incidence of fatigue-related accidents. There has been a plethora of enquiries and reports since the mid 1990s across many jurisdictions and industries, identifying the increased level of fatigue-related risk associated with greater work intensification and the declining efficacy of labor contract negotiations to manage fatigue using prescriptive rule sets (see for example\(^{[24]}\)). The reasons for this are complex and reflect a combination of factors.

First, the variety of WTAs has proliferated considerably over the last 30 yr\(^{[25]}\). This, at least in the post-Robens\(^{[26]}\)
English-speaking world (see “Risk management-based approaches” below), reflects a shift from a highly centralized labor negotiations system to a more de-centralized approach where WTAs are often negotiated at the individual firm level.

Second, there is an increasing awareness that compliance with the prescriptive rules defining WTAs has little or no effect on fatigue attributable to behavior outside the workplace. Non-work related causes of fatigue are known to be a significant source of fatigue-related risk \(^{27}\). For example, care responsibilities, along with family and social commitments, can reduce the amount of sleep obtained in what is, ostensibly, an adequate sleep opportunity between shifts. The same is true for secondary employment, and for commutes to and from the workplace.

Third, for reasons having to do with difficulties staffing shifts, around-the-clock operations, productivity requirements, and/or employees’ earning potential, prescriptive rule sets often allow, or are simply silent about, overtime work. Thus, in many operational settings, employees may find themselves skipping breaks, working extended hours, or scheduling back-to-back shifts. These practices may result in a build-up of fatigue beyond what the prescriptive rules would nominally allow and reduce the opportunity to obtain sleep between shifts.

Fourth, there has been a tendency to assume that compliance with a designated rule set implies a safe system of work and, as a corollary, that non-compliance is unsafe (Fig. 1, top). In reality, there is often poor congruence between safety and compliance with regard to rule sets defining the WTAs (Fig. 1, bottom). For example, deregulation with the goal of improving productivity often results in the potential to create compliant but unsafe WTAs. On the other hand, rule sets for WTAs often create arbitrary “step” thresholds where minor variations in shift and break durations designed to improve productivity are precluded for technical non-compliance and the subsequent inference that the WTAs are now unsafe.

Because fatigue-related accidents are infrequently reported and underestimated\(^{28}\), it can generate a false sense of safety. In smaller organizations that do not have an elaborate safety and reporting infrastructure, such accidents, in general, are even less likely to be reported. However, when compliance with prescriptive rule sets does not necessarily correlate with reduced fatigue-related risk and increased safety, trust in the prescriptive rules erodes and, in turn, compliance may be diminished as well.

**Complex Relationship between Working Time and Fatigue**

The evolution of features in prescription-based WTAs has been implicitly framed by the historical precedents related to the management of physical fatigue. Thus, prescriptive rule sets for fatigue were initially developed to manage physical fatigue and later “appropriated” to manage mental fatigue or sleepiness. Use of the term fatigue to cover both physical and mental fatigue has been a fundamental source of confusion for over a century. Indeed, it is tempting to speculate the extent to which the ambiguities of language might have contributed to, if not caused, many of the ambiguities around WTAs. If mental fatigue had been called sleepiness (as has been noted consistently by Johns\(^{29, 30}\)), the temptation to use regulatory levers originally designed for physical fatigue may well have been avoided.

Prescriptive rule sets for physical fatigue are based primarily on restricting the duration and sequence length of work shifts to limit the accumulation of fatigue. Two underlying assumptions of such prescriptive rule sets are that (1) by specifying maxima on shift duration, the build-up of fatigue is limited; and (2) by specifying minima on time off between duty periods, time for recuperation is protected. However, for mental fatigue in shift work settings, these assumptions are, generally speaking, incorrect. In the case of mental fatigue, therefore, this prescriptive approach may oversimplify the issue.
A key problem is that the relationship between working time and fatigue is non-monotonic. That is, longer working times do not necessarily result in greater fatigue. The reason for this is that two biological processes driving fatigue are involved—one related to time awake (the so-called “homeostatic” process), and the other related to time of day (the so-called “circadian” process). During daytime operations, the two processes are aligned and offset one another, yielding a stable level of alertness through the day. During nighttime operations, however, the two processes are misaligned and effectively amplify each other in building up fatigue. Thus, the relationship between shift duration and fatigue is critically dependent on time of day. This is particularly noticeable in around-the-clock operations, as illustrated in Fig. 2.

Complicating matters further, the effectiveness of breaks between shifts for recuperation is dependent on the presence of physiological sleep. Thus, breaks that do not provide an opportunity to sleep (e.g., because there are other work-related or personal tasks to tend to or no access to a quiet rest area) are generally less effective at reducing mental fatigue than are breaks that allow for sleeping or napping. Furthermore, the effectiveness of breaks between shifts for recuperation is dependent on time of day.
as well. During the so-called “wake maintenance zone” in the late afternoon\(^{36}\), the circadian drive for wakefulness is so strong that it is difficult to maintain sleep. Consequently, night shift workers usually experience sleep restriction even when the duration of their off-duty break allows sufficient time to sleep\(^{37}\). Moreover, during breaks between shift sequences, biological and social forces move the sleep period back to the night\(^{39, 37}\), which prevents adaptation to a daytime sleep pattern and tends to diminish the effectiveness of sleep recuperation further.

Taken together, these complications limit the effectiveness of prescriptive rule sets for the management of fatigue-related risk, especially in working times outside the normal “9 to 5” daytime schedule, and favor the use of risk management-based approaches. This issue has become both prominent and pressing as around-the-clock operations increased after World War II and continued to increase over the past few decades as services and information technology became available 24/7\(^{38}\).

Over the last 20 yr, the drive for greater productivity and an emerging dissatisfaction with the applicability of traditional rule sets has resulted in a search for alternative approaches that are better suited to a more diverse range of WTAs, workers, and tasks\(^{39–41}\).

**Risk Management-based Approaches**

One of the most interesting historical aspects of the regulation of WTAs is the way in which prescriptive rule sets developed and then gave way to risk management-based approaches. This shift can be traced back to the Robens Report of 1972\(^{42}\), in which Lord Robens recommended to the British parliament that the “firm” was the most appropriate governance vehicle for managing safety risk and that legislation should shift from specifying criteria with which the firm should comply to specifying the outcomes that are to be delivered—so called “performance-based legislation”. This was a critical shift in legislative and regulatory focus since legislation now required a firm to ensure a safe system of work but left the firm to determine how best to achieve that end.

While this represents a fundamental change in the way WTAs are regulated and “approved,” performance- and risk-based approaches are relatively novel and, for the most part, lack a significant body of data supporting their efficacy. Nevertheless, over the last few decades, fatigue attributable to WTAs has become a clearly designated workplace safety\(^{43, 44}\) and health\(^{45}\) hazard in almost all developed economies. High-profile accidents directly attributable to fatigue have brought community attention to the risks associated with fatigue-related impairment\(^{46, 47}\).

Significant community campaigns, especially in the road transport sector, along with numerous government reports and inquiries, have identified fatigue as one of the top three workplace safety risks—and arguably one of the most modifiable risk factors influencing workplace safety today\(^{48, 49}\). This trend is further amplified by growing recognition of the extraordinary, but preventable, economic cost of fatigue-related incidents and accidents\(^{50, 51}\).

The appeal of performance- and risk-based approaches lies in their potential to provide a more flexible and nuanced approach to managing fatigue-related risk. The fundamental issue with prescriptive approaches is that they draw relatively arbitrary thresholds for certain features of WTAs (e.g., shift or break durations) and dichotomize the perceived risk of activities below and above that threshold. In reality, the risk of impairment due to fatigue varies on a spectrum—from moment to moment\(^{52}\), individual to individual\(^{53}\), task to task\(^{54}\), and circumstance to circumstance\(^{55}\). As such, the idea of dichotomized thresholds is counterintuitive from a risk-based perspective. Also, as discussed above, dichotomization fails when the relationship between inputs (time worked, time off) and outputs (fatigue, risk) is non-monotonic. Risk-based approaches can enable a more nuanced approach which recognizes that fatigue-related impairment is a spectrum phenomenon that has a complex relationship to the specifics of WTAs.

These issues are formally recognized and addressed in Fatigue Risk Management Systems (FRMS). Over the last decade, a variety of regulatory frameworks have been developed to describe the key elements of FRMS. These regulatory frameworks share some key elements. Most of them sit within broader regulatory standards for risk and safety management. For example, many make reference to ISO31000/AS4360 standards for risk management (RM) and/or one of several standards for Safety Management Systems (SMS) (e.g., AS 4804/4801). In this context, compliance to a simple prescriptive rule set is replaced by a process in which the risk associated with proposed WTAs is assessed using a standardized methodology and must be demonstrably mitigated to a level considered acceptable. This so-called “safety case” may, in some cases, then be accepted (approved) by a regulatory body.

This risk-based approach gives dimension to risk as a product of the *likelihood* and the *consequence* of a fatigue-related error\(^{56}\). Each of these two factors can be divided into five ordinal levels in a semi-quantitative manner, recombined using the matrix in Fig. 3, and then allocated...
to one of four ordinal risk categories (Low, Medium, High, or Very High). The greater the assessed risk, the greater the level of risk mitigation required.

FRMS policy frameworks have been developed for a variety of industries and even specific organizations\(^58\text–}^{63}\). There is some degree of variation in how these frameworks define and describe FRMS and how risk is measured, mitigated, and monitored. There is, however, a common pattern to these of policy frameworks, training and education, risk assessment and mitigation, and monitoring and review. The following sections describe these elements in more detail.

### Policy Framework

Most FRMS policy frameworks address a range of governance issues, including the legal defensibility of the WTAs in the event of a fatigue-related accident or injury. This can be especially critical in essential services industries where rigidly prescribed working hours can be operationally difficult to comply with. In many industries, such as emergency services, health care, and law enforcement, WTAs are often extended because the withdrawal of services is socially or politically intolerable. This has caused difficulties for organizations that extended working hours to meet community expectations but were subject to post-hoc litigation when fatigue-related accidents occurred.

Risk-based approaches provide a plausible solution to this dilemma. Provided the risks of withdrawing the service (due to exceeding the prescribed WTAs) are greater than the risks associated with continuing to work, the decision to extend working hours should be legally defensible. In the case of industries where the withdrawal of services carries significant immediate consequences (e.g., withdrawal of health care or emergency services), assuming reasonable measures have been taken to provide resources and staffing that should normally be adequate, the decision to extend the WTAs is readily defensible.

On the other hand, in more clearly commercial settings such as transportation or manufacturing, it may be difficult to justify decisions to extend the WTAs. This is because the primary beneficiary of the extension is the employer (improved productivity) and sometimes the employee (increased income), while the risk may be carried by others (e.g., other road users or the community). Using a risk-based criterion enables organizations to define the benefits (or lack thereof) to extended WTAs, and weigh them against the risks, in a legally defensible manner.

Prescriptive rule sets for WTAs often fail to assign responsibility for fatigue-related risk in a comprehensive...
manner. They tend to assume that responsibility for fatigue falls primarily to the employer via the WTAs. Research shows that the behavior of employees outside of the workplace (and WTAs) can heavily influence subsequent fatigue-related risk in the workplace. In FRMS, there is no obvious, default assignment of responsibility. Ideally, there needs to be a clear articulation of the roles and responsibilities of employees with respect to the FRMS. Importantly, risk-based policy enables organizations to introduce a “shared responsibility” framework in which the employer is primarily responsible for work-related causes of fatigue and the employees for non-work-related causes of fatigue.

A corollary of the “shared responsibility” model is the requirement that employees report being unfit for work due to insufficient sleep between shifts. While often overlooked, many national workplace health and safety (WHS) jurisdictions have dual duty-of-care obligations. From a risk-based perspective, it is not unreasonable to require employees to report being unfit for work as part of their “shared responsibility” obligations. In recent years, many jurisdictions have seen a significant increase in personal responsibility clauses in fatigue risk management policies. Note that this also presumes a level of predictability in the rostering of employees, so that they can plan ahead and adapt their off-duty time to meet their personal responsibility. In industries with unpredictable work hours, on-call duty, and/or “on the fly” changes in work schedules, this may be difficult.

There are obvious cultural barriers (e.g., potential social stigma) to encouraging employees to report being unfit for work. However, safety culture initiatives that emphasize “just culture” frameworks, which are based on a collective understanding of the difference between blameworthy and blameless actions and situations, are likely to help promote self- and/or peer-reporting of fatigue-related impairment. Similarly, a more sophisticated understanding of how authority gradients can influence the reporting process, along with how the language used in these reports can minimize blame or stigma, are important factors shaping the identification and management of individual risk events.

Finally, where individual risk events are managed through a self- and/or peer-reporting framework, those identified as “at risk” should reasonably expect the organization to respond by managing the event in a fair, just, and unambiguous way. The “just culture” literature provides guidance on how best to implement this in the context of FRMS.

Training and Education

There is a long history of training and education with respect to fatigue-related risk. Since the mid-1980s, a variety of providers have developed training materials for delivery in workplaces, although they are often outdated or fail to address relevant issues. To date, the primary purpose of these training programs has been to identify the “3 C’s”:

- **Consequences** of shift work (e.g., increased incidence of accidents and injury, increased incidence of lifestyle-related illnesses, and increased incidence of psychosocial dysfunction);
- **Causes** of these conditions (e.g., the cognitive and biological effects of fatigue due to reduced sleep opportunity, circadian misalignment, and increased social isolation from family and friends);
- **Compensatory** changes to minimize the potential harm associated with shift work (e.g., improved sleep hygiene, better lifestyle choices, reporting fatigue).

Because prescriptive approaches use rule sets as the principal means to control risk, the associated training and education often places less emphasis on employer responsibilities relative to employee responsibilities. Traditional training and education programs have typically been employee-focused, concerned primarily with identifying and mitigating non-work causes of fatigue. In a risk-based framework, factors other than the “3 C’s” outlined above can be readily included. Risk-based approaches typically include training and education with respect to:

- Understanding personal and organizational roles and responsibilities with respect to managing fatigue-related risk;
- Competency in:
  - assessing the risks associated with WTAs using standardized methodologies,
  - identifying and implementing evidence-based risk mitigation strategies at the personal, work group, and organizational level;
- Monitoring, reviewing, and revising fatigue management practices using evidence-based performance metrics.

Some jurisdictions have developed more formalized risk-based training curricula. For example, in Australia, there are now standardized curricula for fatigue risk management defined within the vocational educational sector. These curricula are competency-based and are divided into two categories:

- Employee focused: these curricula include the typical
materials included in traditional prescriptive training (the “3 C’s”), including strategies to improve sleep hygiene, reduce social isolation, and improve lifestyle choices73).

- Supervisor focused: these curricula add materials necessary to help staff responsible for determining the hours that staff work and provide background materials that help supervisors and managers identify and mitigate factors that increase the risk of fatigue-related accidents and injuries74).

More recently, in higher education, undergraduate and post-graduate course material on fatigue risk management has been introduced75). These subjects typically sit within broader human factors or WHS/safety management programs and provide students with the skills necessary to design, implement, and evaluate FRMS.

**Risk Assessment and Mitigation**

Perhaps the most fundamental change in transitioning from a prescriptive to a risk-based approach for managing fatigue has been the move toward quantifying the risk associated with specific WTAs. This significant change in practice has not been without controversy, as it is predicated on the ability of practitioners to assess the relative risk associated with different WTAs in a valid and reliable manner.

Assessing the relative risk is complicated by the validity of what is measured as a proxy measure or leading indicator (early predictor) of the likelihood of fatigue risk. Based on the work of James Reason76), leading indicators can be placed along an event trajectory based on their causal proximity to a fatigue-related incident (Fig. 4). Using this event trajectory, five categories of indicators have been identified:

1. Reductions in sleep opportunity which make it difficult to obtain sufficient sleep to work safely
2. Reductions in obtained sleep that are inconsistent with working safely
3. Signs and symptoms of fatigue that are inconsistent with working safely
4. Task-related errors consistent with a fatigue-related error
5. Actual fatigue-related accidents and near-miss incidents

Dawson and McCulloch77) outlined the first integrated approach for managing fatigue using these categories, which was subsequently incorporated in a range of FRMS guidance materials. The five categories involve the following risk assessments and mitigation methodologies:

<table>
<thead>
<tr>
<th>Latent Errors</th>
<th>Active Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep opportunity</td>
<td>1. Rules of rostering</td>
</tr>
<tr>
<td>Obtained sleep</td>
<td>2. Prior Sleep/Wake data</td>
</tr>
<tr>
<td>Behavioral Symptoms</td>
<td>3. Symptom checklists</td>
</tr>
<tr>
<td>Behavioral Symptoms</td>
<td>Self-report behavioral scales</td>
</tr>
<tr>
<td>Behavioral Symptoms</td>
<td>Physiological monitoring</td>
</tr>
<tr>
<td>Fatigue-related errors</td>
<td>4. Fatigue-prone strategies</td>
</tr>
<tr>
<td>Fatigue-related errors</td>
<td>Error analysis system</td>
</tr>
<tr>
<td>Fatigue-related incidents</td>
<td>5. Incident analysis system</td>
</tr>
</tbody>
</table>

Fig. 4. Event trajectory of leading indicators of fatigue-related incidents (modified from Dawson and McCulloch77).

1. Sleep opportunity
2. Obtained sleep
3. Behavioral indicators of sleep
4. Threat and error management
5. Incident analysis

See Dawson and McCulloch77) for an in-depth discussion of these categories.

A complementary approach to assessing the relative risk associated with different WTAs is the use of validated mathematical models of fatigue78, 79). Such models convert the timing and duration of sleep opportunities or actually observed sleep associated with work schedules into predictions of fatigue risk, based on equations for the known neurobiology of fatigue (the “homeostatic” and “circadian” processes discussed above)50). From these fatigue risk predictions, a measure of “risk exposure” can be derived as the product of the magnitude of predicted fatigue risk and the amount of time passing (i.e., the area under the curve). The relative risk associated with different WTAs can be determined by comparing the corresponding predictions for risk exposure. Procedures for applying mathematical models to compare schedules in terms of risk exposure are beyond the scope of this paper, but examples can be found in the literature51, 80, 81).

This mathematical modeling approach has been validated against observed accident risk in US-based railroad operations51). However, its validity may or may not extend to other operations, based primarily on the extent to which employees make use of WTA-based sleep opportunities in a predictable manner. In international air travel, for example, it turns out to be quite challenging to predict pilots’ use of sleep opportunities in a generalizable manner82). It should also be noted that the use of mathematical models of fatigue in the context of FRMS is appropriate only for relative comparisons among WTAs; these models have not
been validated to provide absolute assessments of fatigue risk as they do not account for differences in risk associated with different outcome metrics, job tasks, individuals, or circumstances\(^\text{50}\). In this context, mathematical models of fatigue can be particularly useful to help guard against work schedules that would be permitted under prevailing prescriptive rules but are relatively unsafe. This is straightforward to implement in industries where work schedules are constructed by automation or optimization, such in commercial aviation, but it can also be effective in manual schedule construction. Based on model predictions, the overall risk of the WTAs can be estimated and ‘warnings’ can be issued without these necessarily representing a hard limit on acceptable work schedules.

**Monitoring and Review**

In operational settings where safety is of primary concern, one of the most critical changes associated with moving from a prescriptive, compliance-based system to a risk-based system is the need to embed FRMS within the overarching SMS. This means that the FRMS, along with the broader SMS, should be monitored to provide evidence to support its efficacy. In general, this requires a set of performance metrics that enables the organization to determine whether their FRMS are (a) operating as intended and (b) managing the risk effectively.

A monitoring and review system will require metrics to evidence the efficacy of the following:

- The policy framework: Has the organization ensured everyone has been informed of their roles and responsibilities with respect to the FRMS? Does everyone know the policies and procedures and the governance structures though which the FRMS will operate? Have all contractors been covered in this relationship?

- Training and education: Has everyone undergone the appropriate levels of training? Is there evidence of demonstrated competency in assessing and controlling fatigue-related risk?

- Risk assessment and mitigation: Does the risk assessment methodology provide data that enables staff to determine its efficacy? Are the risk controls effective? Have new risk factors emerged or are there other unintended consequences from the FRMS framework? In particular, what do the (a) sleep opportunity, (b) obtained sleep, (c) signs and symptoms, and (d) error data say about the incidence of fatigue and the level of safety in the workplace and/or the efficacy of controls? Based on the work of Reason\(^\text{76}\), a good SMS (and by inference good FRMS) should shift the focus from measuring only low frequency/high consequence events (i.e., fatigue-related accidents) to including higher frequency/ lower consequence metrics (e.g., fatigue-related errors). That is, the incidence of metrics linked to “Risk assessment and mitigation” items listed above (a)–(d) should all be considered, since significant declines in their rate of occurrence are easier to detect statistically and are likely to be broadly indicative of the actual subsequent incident rate—although further research is needed to substantiate this for specific industries\(^\text{33}\).

Because it can be challenging to measure risk and the associated leading indicators in operational settings, organizations often simply measure compliance with the procedures outlined in the FRMS framework rather than determine actual fatigue and risk-based outcomes. Conceptualized this way, FRMS can be seen as involving predictive (prospective), proactive (day of operations), and reactive (retrospective) procedures designed to help prevent or mitigate fatigue and risks associated with fatigue\(^\text{84}\). In this framework—aided by FRMS tools such as predictive mathematical modeling of fatigue for scheduling, proactive fatigue monitoring in the workplace, and reactive post-incident follow-up—the regulatory focus is not on fatigue risk per se, but rather on procedural compliance under the assumptions that (a) compliance can be assessed reliably; and (b) compliance is congruent with safety. This approach puts a premium on effective strategies for detecting “active errors” (fatigue-related errors and incidents; Fig. 4) to detect and fix any break-downs in the relationship between (presumed) compliance and safety—lest the FRMS becomes ineffective (Fig. 1).

An FRMS approach that includes measurements of actual fatigue risk would be expected to be superior in its effectiveness to curb the risk. However, the development of effective and reliable tools for detecting fatigue and fatigue risk is still an area of ongoing research\(^\text{70}\), and monitoring and review strategies based on leading indicators and/or procedural compliance will likely remain the norm until such tools become more widely available. It is worth noting, though, that some relatively simple approaches for the assessment of fatigue have been deployed. For example, Air New Zealand used a subjective fatigue scale (the Samn-Perelli scale\(^\text{85}\)) entered into the board computer during flight (before landing at “top of descent”) for many years. Acknowledging the obvious problem of potentially accurate reporting, such an approach can nonetheless be useful when the fatigue assessments are made outside the context of an adverse event. These kinds of initiatives
are predicated on a relatively mature safety culture for the organization, however, and therefore typically remain sponsored and supported by companies and employees rather than part of a regulatory requirement.

**Conclusion**

Prescriptive rule sets have been the starting point of a journey in the management of fatigue-related risk that is still unfolding. There is little doubt that FRMS represent a significant positive development, especially for safety-sensitive, around-the-clock operations that are particularly sensitive to mental fatigue. Major policy frameworks for FRMS have been disseminated by a variety of regulators in a cross-section of industries, especially in the English-speaking world. Not surprisingly, the most salient examples first emerged in the capital-intensive end of the transportation sector—aviation and railroad transportation. In these industries, economic advantages, significant investment in safety infrastructure, and a long tradition of using a SMS approach have made the transition to FRMS relatively straightforward. In these industries, FRMS have been enthusiastically embraced as a way to resolve the obvious paradoxes related to safety and productivity associated with more traditional, prescriptive approaches to controlling fatigue-related risk. On the other hand, the transition to FRMS has been less enthusiastically embraced in industries characterized by a large number of small operators (e.g., road transportation), where labor negotiations have been traditionally adversarial (e.g., mining), or in industries where aspects of the safety culture are perhaps less mature (e.g., health care).

Despite the logical appeal of a risk-based approach, FRMS constitutes a paradigm shift in the regulation of WTAs that has been implemented in a largely untested manner in many jurisdictions. To a certain extent this is unavoidable—the efficacy of FRMS cannot be assessed *a priori*. In practice, the emergence of FRMS has piggy-backed on the success of the broader evidence-backed SMS movement\(^{86–89}\). It has generally been accepted that SMS approaches are to be preferred over prescriptive safety rules because of reduced regulatory costs and because they allow some operational flexibility to promote safety (and efficiency) above and beyond what prescriptive rules can offer. SMS approaches are also more likely to succeed than prescriptive ones, given the greater external enforcement costs of the latter\(^{90}\). These advantages are assumed (reasonably) to extend to FRMS.

It is important to note that the success of a risk-based approach is predicated on a base level of safety culture, infrastructure, and risk management expertise within the implementing organization and its relevant regulatory agencies. While this may be a realistic assumption for larger organizations and regulatory agencies in larger countries (those with significant pre-existing safety infrastructure and resources), smaller organizations may well struggle with the greater sophistication required to implement a full-fledged FRMS as compared to prescriptive rule sets. Indeed, recent FRMS initiatives in Australia in the road transportation sector were difficult to implement for smaller operators because they often lacked the resources and skills necessary to implement and monitor the more complex requirements of FRMS, and regulators were often under-resourced and inadequately skilled to embrace an evidence- or performance-based safety culture\(^{91}\). In these settings, smaller enterprises might be supported in the development of industry ‘templates’ or simplified versions of FRMS developed by larger organizations.

There is little doubt that poorly implemented or monitored FRMS may result in the deregulation of an organization’s WTAs. This would also lead to a corresponding increase in the potential for fatigue-related risk. Importantly, a shift to an FRMS-based approach changes the demands on regulators charged with monitoring and reviewing risk- and performance-based systems. Compliance with prescriptive rule sets is typically unambiguously determinable, whereas compliance within a framework based on fatigue and risk (concepts that cannot readily be measured directly) may be more difficult to assess, given that mere compliance with the procedures laid out in the FRMS policy framework does not guarantee the achievement of desired risk outcomes. Recent experience in the United States and Australia has also suggested that, for regulatory agencies with a long history of compliance-based audit and review, SMS approaches in general and FRMS approaches in particular can prove challenging to regulate effectively when regulatory oversight is limited and sanctions difficult to enforce\(^{88}\).

Given the resources and safety culture required to implement FRMS effectively, it may not necessarily be a better choice for managing fatigue-related risk. It is also probably fair to say that the application of risk-based approaches has been limited to the safety implications of WTAs, with little consideration of other implications such as the psychosocial and health-related impacts\(^{92, 93}\). While the ability to quantify and predict fatigue-related risk has advanced considerably over the last two decades, the ability to assess the psychosocial consequences of fatigue is
still underdeveloped. The problem is that health-related consequences of fatigue typically emerge over long time intervals, often extending past retirement\footnote{Transport Canada Work/Rest Rules for Railway Operating Employees. https://www.tc.gc.ca/eng/railsafety/rules-tco140-364.htm. Accessed August 2, 2017.}. The ability to address these facts of fatigue is often beyond the currently recognized reach of regulatory agencies – even in the most advanced economies\footnote{Menckenh HL (1920) Prejudices: second series, Chapter 4: The Divine Afflatus, 158, Alfred A. Knopf, New York.}

Future evolutions of the FRMS regulatory framework may try to move further toward a hybridized approach that permits organizations and regulatory agencies to optimize the benefits of both risk-based and prescriptive approaches. This may permit small firms, for whom the cost-benefit analysis does not support the introduction of FRMS, to keep using a low-cost, low-complexity prescriptive system. On the other hand, for those organizations where the cost-benefit analysis warrants a significant investment, the FRMS may prove to be a preferable option. Indeed, this hybrid model already exists in some jurisdictions where FRMS form either (a) the basis for the “safety case” supporting an organization exceeding prescriptive WTAs (e.g., flight and duty times in US commercial aviation; road transport in Australia or rail transport in New South Wales), or (b) the basis for risk assessment of different WTAs that fall within an “outer limits” prescriptive framework (e.g., the US rail industry, which uses fatigue modeling to determine the relative risk of different “legal” WTAs).

In the long run, risk-based approaches also carry significant potential to correct market anomalies in a range of industries and provide better models for internalizing the currently externalized costs associated with fatigue-related accidents. For example, non-compliance with prescriptive regulations is often economically “rational” since the profits from greater productivity are “privatized” and the costs associated with accidents are “socialized” (i.e., paid for by the rest of the industry via insurance premiums or by the community via taxation). Risk-based approaches that more accurately quantify the risk associated with WTAs could be used to set insurance premiums (and also expose excess payments) in a way that reduces the opportunities to invisibly externalize costs. Moreover, if insurance companies could demonstrate that a company has violated their FRMS, then they could reasonably refuse to pay out on insurance policies. For industries where the risk of regulatory enforcement has historically been so low as to incentivize non-compliance, this would provide a clear pricing signal to the market to encourage compliance and promote safe practices.

Acknowledgements

The authors thank Tomas Klemets and an anonymous reviewer for helpful feedback on an earlier version of this paper. The authors also thank the expert panel for their thoughtful review of the consensus statements.

References


78) Mallis MM, Mejdal S, Nguyen TT, Dinges DF (2004) Summary of the key features of seven biomathematical


