The Evaluation of Fatigue as a Performance Shaping Factor in the Petro-HRA Method

Martin Rasmussen* & Karin Laumann

Department of Psychology, Norwegian University of Science and Technology

NO-7491 Trondheim

Norway

Available online 26 June 2018 in Reliability Engineering and System Safety

https://doi.org/10.1016/j.ress.2018.06.015

*Corresponding author; martin.rasmussen@ntnu.no

Abstract: In the development of the Petro-HRA method (Bye et al., 2017), a human reliability analysis (HRA) developed for the petroleum industry, a number of factors believed to effect human performance were reviewed and considered for inclusion in the method's performance shaping factor (PSF) taxonomy. The method was created for prospective risk analysis of post-initiator events and it was created as a method that focused on including the most important PSFs, rather than attempting to include all aspects of human performance. This paper assess whether fatigue should be among the PSFs included. This article presents: (1) how fatigue is included in current human reliability methods; (2) fatigue and its underlying aspects; (3) how these aspects affect performance and; (4) the consideration of including fatigue as a PSF in Petro-HRA. Four possible PSFs based on the causes of are suggested: Sleep deprivation, Shift-length, Non-day shift, and Prolonged task performance. However, due to the relative low impacts of the PSFs and the Petro-HRA's focus on only the strongest PSFs, the final method did not include any of the suggested fatigue PSFs.

Keywords: Human Reliability Analysis, Fatigue, Petro-HRA, Performance Shaping Factors.

1. Introduction

Fatigue has been listed as one of the contributing factors in several incidents where human error played an important role, such as the Chernobyl and the Challenger Space Shuttle accident, and the near accident at Three Mile Island (Mitler et al., 1988), indicating that fatigue could be a significant influence in human performance. In a high-risk industry, several measures are taken to identify potential risk and reduce it. One of the measures is conducting analyses that focus on the accident sequences following an initiating event (e.g. a loss-of-coolant accident in a nuclear power plant or a loss of well control in offshore petroleum drilling) and estimating the probability that the initiating event will lead to a major accident (often defined as reactor core damage in a nuclear power plant or a loss of lives and/or large environmental damages in the petroleum industry). In the nuclear industry these analysis are called probabilistic safety assessment (PSA) while quantitative risk assessment (QRA) or total risk analysis (TRA) are used in the petroleum industry. There are some differences between a PSA, a QRA and a TRA (Øien & Sklet, 1999), but they have in common that that they include accident sequences that follow an initiating event, that they are predominantly focused on technical issues, and that the human element can be included through a human reliability analysis (HRA). It has been suggested that fatigue is not sufficiently been prioritized by current HRA methods, and that its inclusion should be increased in future methods (Griffith & Mahadevan, 2011).

An HRA is defined as "any method by which human reliability is estimated" (Swain, 1990, p. 301) and generally consists of three steps: (1) identifying possible human errors and contributors, (2) model human error, and (3) quantify human error probabilities (HEPs), with human error reduction at times included as fourth step (Kirwan, 1994). The nuclear industry has a long tradition of including HRA as part of their larger technical safety assessments, and has been the leading industry both in terms of use and development of HRA (see Boring, 2012; Rasmussen, 2016). This has not been the case in the petroleum industry where the QRA or TRA generally has not included an HRA (Skogdalen & Vinnem, 2011; Vinnem, 1998; Øien & Sklet 1999). In the years since the 2010 Deepwater Horizon blowout accident (Graham et al., 2011), the HRA activity in the petroleum industry has increased with analyses being conducted using methods from the nuclear industry (van de Merwe, Hogenboom, Rasmussen, Laumann, & Gould, 2014) and the development of a method adapted to the petroleum industry. Petro-HRA (Bye et al., 2017) is a complete HRA method including steps on scenario definition, qualitative data collection, task analysis, human error identification, human error modelling, human error quantification and human error reduction. The method is intended for prospective analysis of post-initiator scenarios in the petroleum industry. Several HRA methods from the nuclear industry inspired parts of the method with the strongest link being to the Standardized Plant Analysis of Risk – Human Reliability Analysis (SPAR-H; Gertman, Blackman, Marble, Byers, & Smith, 2005), a method that was identified as the most applicable method for the petroleum industry in a review (Gould, Ringstad & van de Merwe, 2012) conducted prior to the Petro-HRA project (Laumann et al., 2014). SPAR-H was originally intended for retrospective analysis (Boring, 2015), while both actual use and the final 2005 guideline include both prospective and retrospective use of the method. The quantification framework in Petro-HRA where a nominal HEP is multiplied with performance shaping factors (PSFs) to provide an actual HEP, is adapted from SPAR-H. In the Petro-HRA development the PSF taxonomy used in SPAR-H was revised based on relevance to the petroleum industry, analysts' experiences with SPAR-H and psychological research. Many factors were reviewed and considered (Laumann & Rasmussen, 2016; 2017; Rasmussen, 2016; Rasmussen & Laumann, in review; Rasmussen, Standal & Laumann, 2015) based on the four following questions (Rasmussen, 2016, p. 6):

1. Can it have a significant effect on human performance in an accident scenario?

2. Can it be, and if so, how should it be measured and evaluated?
3. Can it be, and if so, how should it be quantified?
4. Can it add to the method without overlapping with any other PSF?

The final PSF taxonomy of the Petro-HRA method included nine PSFs; Time, Threat stress, Task complexity, Experience/Training, Procedures, Human-machine interface, Physical working environment, Teamwork and last PSF which includes Attitudes to safety, work and management support.

This paper presents the review of the fatigue factor, one of the reviewed factors that were not included in the final PSF taxonomy. The paper starts with a presentation of how fatigue is included in current HRA methods, followed by a discussion of the different aspects of fatigue, how they could be included in an HRA method, and the discussion whether to include fatigue in the Petro-HRA method. The inclusion of fatigue in HRA methods has been explored in a review by Griffith and Mahadevan (2011); however, they limited their definition of fatigue to only include sleep deprivation. The choice of not including a Fatigue PSF in Petro-HRA has been mentioned in previous papers (Bye et al., 2017; Laumann & Rasmussen, 2016), however without the space to go into detail on the process. The authors believe that this paper will contribute to transparency in the Petro-HRA method development and potentially serve as an input to future HRA methods.

1.1 Fatigue in Current HRA Methods

As research on the connection between fatigue and human performance has existed for some time, at least since Thorndike's (1900) famous studies, it would seem likely that risk associated with fatigue had been quantified and included in methods created to estimate human performance or reliability. This section evaluates how, if at all, current HRA methods include fatigue.

1.1.1 SPAR-H

The SPAR-H method (Gertman et al., 2005) is an HRA method that estimates HEP through multiplying one of two nominal HEPs with PSFs. The two nominal HEPs are 0.01 for diagnostic tasks and 0.001 for action tasks. Diagnostic tasks are further elaborated to mean not only diagnosis, per se, but rather the entire spectrum of cognitive processing including aspects such as interpreting information, understanding a scenario and decision-making (Whaley et al., 2011). PSFs represent aspects of individual characteristics, the environment, the organization, or the task that decrements or improves human performance, thereby increasing or decreasing the likelihood of human error (Boring & Blackman, 2007). The eight PSFs included in the SPAR-H method are: *Available Time*, *Stress/Stressors*, *Complexity*, *Experience/Training*, *Procedures*, *Ergonomics/HMI*, *Fitness for Duty*, and *Work Processes* (Gertman et al., 2005). Each of the PSFs in SPAR-H has a set of levels – such as *nominal time* or *highly complex* – with a description and a multiplier. The SPAR-H method does not include a separate PSF for fatigue; however, it is included in the *degraded fitness* level of the *Fitness for Duty* PSF described in the SPAR-H manual:

> The individual is able to carry out the tasks, although performance is negatively affected. Mental and physical performance can be affected if an individual is ill, such as having a fever. Individuals can also exhibit degraded performance if they are inappropriately overconfident in their abilities to perform. Other examples of degraded fitness include experiencing fatigue from long duty hours; taking cold medicine that leaves the individual drowsy and nonalert; or being distracted by personal bad news (such as news of a terminal illness diagnosis of a loved one). (Gertman et al., 2005, pp. 25–26)

The *degraded fitness* level has a multiplier of five leading to a HEP of 0.05 for diagnostic tasks and 0.005 for action tasks if all other PSFs are nominal. Fatigue, however, is only a small aspect of this PSF, and the *Fitness for Duty* PSF is rarely used in prospective analysis. It should be noted that the SPAR-H guideline (Gertman et al. 2005) does emphasize that it is a simplified method and if a detailed analysis of human performance aspects is required a more in-depth analysis such as A Technique for Human Event Analysis (ATHEANA; NUREG-1624, 2000) is recommended (see section 3.2.4). The categorization as a simplified method stems from that SPAR-H was created as a simplification and generalization of the Technique for Human Error Rate Prediction (THERP; Swain & Guttmann, 1983; see section 3.2.2) and the ASEP (Swain, 1987) methods (Boring & Blackman, 2007). One important step in the simplification was the use of a rather small set of PSFs that were found to be the most important in shaping performance.

1.1.2. THERP

THERP was presented in the *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications* (Swain & Guttmann, 1983). THERP uses a set of tables that provides HEPs for a large amount of specific tasks. The HEPs are then modified with multipliers, and dependencies are calculated through the tables provided. The modification through multipliers is similar to the modification PSFs have in methods such as SPAR-H, but THERP has a much higher focus on providing a HEP for the specific task and fewer modifiers. The first chapters of the handbook describe several forms of fatigue as physiological stressors (e.g., sleep deprivation, long work hours, fatigue caused by noise) and their possible effect on performance. However, the handbook states that as it would be impossible to quantify everything that affects human performance, several of the aspects discussed in the first chapters are included in a category called “other PSFs.” These PSFs are not included in the tables and multipliers used for determining HEP in THERP. The analyst is, however, reminded several times in the guidelines to evaluate if the “other PSFs” are affecting performance and use his or her own judgment in determining the effect. Problems of fatigue caused by prolonged tasks are not explicitly referenced in the tables or the PSFs, though a PSF for low taskload is included. This could overlap with the prolonged monotonous tasks. The PSF has a multiplier of two for both skilled and novice personnel (see THERPs Table 20-16; Swain & Guttmann, 1983).

1.1.3. HEART

Human Error Assessment and Reduction Technique (HEART; Williams, 1988; 1992) is an HRA method that estimates HEP through 14 generic task types (GTTs), each with a proposed nominal human unreliability and 38 error-producing conditions (EPCs), each with a maximum multiplier. The EPCs in HEART correspond to the PSFs in SPAR-H, with the exception that they are not graded through levels; instead, the analyst chooses how to weight the EPCs (from 0.1 to 1.0) based on the change they are thought to have in this context (a weight of 1.0 will result in the maximum multiplier). None of the 14 generic task types specifically make a reference to fatigue or aspects of fatigue, although two of the EPCs include aspects of fatigue: "Prolonged inactivity or highly repetitious cycling of half hour low mental workload tasks" (Williams, 1992, p.23) with multipliers of 1.1 for first half hour and 1.05 for each hour thereafter; and "Disruption of normal work-sleep cycles" (Williams, 1992, p.23) with a multiplier of 1.1. In an update of the HEART method, the multiplier for the "Disruption of normal work-sleep cycles" was increased to "1.2 (compounded) for every 24 hours' sleep lost to at least 72 hours without sleep" (Williams & Bell, 2017, p.885). The update also introduced two additional EPCs, including one relevant for fatigue: Time-of-day (from diurnal high arousal to diurnal low arousal), with a maximum multiplier of 2.4. The time around 16:00 hrs is identified as having the highest reliability, 03:00 hrs as the lowest, and a post lunch dip is identified around 14:00 to 15:00 hrs with a multiplier of 1.3 (Williams & Bell, 2017).

1.1.4 ATHEANA

ATHEANA (NUREG-1624, 2000) is based on the premise that "significant human errors occur as a result of 'error-forcing contexts' (EFCs), defined as combinations of plant conditions and other influences that make operator error very likely" (NUREG-1624, 2000, p. xv). The focus of the ATHEANA method is somewhat different than other HRA methods. It is based upon the observation that most HFEs are not due to simply forgetting a step in the procedures or not noticing an alarm, rather "HFEs occur when the operators are placed in an unfamiliar situation where their training and procedures are inadequate or do not apply, or when some other unusual set of circumstances exist" (NUREG-1880, 2007, p. 1-1). Following this idea, the ATHEANA method includes identifying operational vulnerabilities that could set up potential unsafe actions and identifying plausible deviations from normal conditions, in addition to the conventional steps of identifying human actions, defining the HFE, and determining the HEP. ATHEANA does not operate with a fixed set of EFCs or PSFs (although a list of commonly used PSFs is provided), rather it promotes the systematic search for these and the unsafe actions to which they may lead. The additional steps of the ATHEANA method add thoroughness to the method, although systematic searches are time consuming. The SPAR-H manual suggests a solution to this in using SPAR-H (which is a simplified method) for most events while promoting the use of ATHEANA in situations that require a "detailed analysis of the human performance aspects of an event" (Gertman et al., 2005, p. v). Both EFCs and PSFs in ATHEANA could include aspects of fatigue, depending on to what the systematic search leads. *The impact of time of day on operator performance* (NUREG-1624, p. 9-62) is included as an example of a PSF that could be relevant in a deviation scenario. Other fatigue aspects are also likely to be included in a deviation scenario, although the inclusion of any EFC or PSF in ATHEANA will depend on the analyst.

1.1.5 CREAM

Cognitive Reliability and Error Analysis Method (CREAM; Hollnagel, 1998) is an HRA method that estimates the probability of failure through four types of cognitive functions (observation, interpretation, planning, and execution; and a total of 13 generic cognitive function failures connected to these cognitive functions, each with a nominal failure rate and uncertainty bounds) modified though nine common performance conditions (CPC). The generic cognitive function types resemble the GTTs found in HEART, but CREAM differs in that it focuses only on the cognitive aspects. In addition, the CPCs in CREAM resemble the PSFs seen in SPAR-H. An aspect of fatigue can be accounted for through the CPC called *time of day (circadian rhythm)* which is described as “the time at which the task is carried out, in particular whether the person is adjusted to the current time” (Hollnagel, 1998, p. 244), which is accompanied by the levels of *day-time (adjusted)* or *night-time (unadjusted). Day-time (adjusted)* is the nominal level with a multiplier of 1 in all four cognitive functions, while *night-time (unadjusted)* has a multiplier of 1.2 for all four cognitive functions.

2. Method

To find the underlying elements of fatigue, a review of relevant literature was performed. Research articles, review articles, and academic books that examined the effect of fatigue on HEP were found through Boolean searches of PsycNET, PsycINFO, Google Scholar, and ISI Web of Science. Only results that could be translated into multipliers where included, and where reviews existed these were preferred over single studies. The data sources were analyzed using a thematic analysis. Thematic analysis offers a theoretically flexible approach allowing several types of data sources and that can be used for different proposes (Howitt, 2010). Thematic analysis can be used to systematize a dataset into different sub-themes within a larger theme. This suited this paper well as the intention was to systematize the larger theme of fatigue through the aspects or causes of fatigue. The thematic analysis was conducted using the NVIVO 10 software. A thematic analysis often includes several rounds of coding; in this case, however, the different aspects of fatigue were relatively clearly separated by the cause of the fatigue, leading to a single round of coding. All data sources were coded by the cause(s) of fatigue upon which they focused. The use of NVIVO 10 and thematic analysis ensured a systematic approach to coding the data. Following the thematic analysis with a meta-analytical approach was considered, but not employed as recent meta-analyses existed for the different fatigue aspects identified.

3. Results

While fatigue is a concept that is well established in our everyday vocabulary, it is a concept that lacks an agreed scientific definition (Noy et al., 2011; Williamson et al., 2011). The difficulty in agreeing on a definition stems from the multifaceted nature of the fatigue concept. Fatigue has a wide range of causes, experiencing fatigue can lead to a wide range of effects that can be measured in several different ways, and the duration and alleviation will differ based on these variables and individual differences. In this article we have chosen to use Williamson et al.'s (2011, p. 499) definition of fatigue: "A biological drive for recuperative rest." This definition was chosen as it defines fatigue at a high-level, thereby including several possible causes, effects and alleviations.

Our thematic analysis found four distinct aspects of fatigue categorized by cause: sleep deprivation, shift length, non-day shift, and prolonged task performance.

3.1 Sleep Deprivation

Sleep deprivation refers to the extent of time since sleep or a reduction in the quality or quantity of sleep. Griffith and Mahadevan (2011) reviewed incidents where sleep deprivation had been found to be a contributing factor and studied the effect of sleep deprivation on performance. Their review presents several serious incidents, such as Three Mile Island, Chernobyl, and Exxon Valdez, where the investigation reports have cited sleep deprivation as a partial cause. They also discussed several incidents of employees, including operators, sleeping while on duty in high-risk industries.

One way of presenting the effects of sleep deprivation is to compare it to the effects of alcohol consumption which has similar symptoms such as slower responses, failures in attention, and failures in suppressing inappropriate strategies (choosing simpler, but riskier strategies) (Mitler et al., 1988). This has been a frequent comparison in the transport industry where the effect of alcohol is a well-researched topic. However, it should be noted that while there are similarities between the effects of sleep deprivation and alcohol consumption the underlying physiological and psychological effects are different leading to some limitations in the comparison. Six studies reviewed by Griffith and Mahadevan (2011) showed that performance after 17–25.1 hours of wakefulness had the same effect on performance as a blood alcohol percentage of 0.05–0.10, values above the legal limits of driving in most countries. In a continuation of this work Griffith and Mahadevan (2015) used a meta-analytical approach to derive PSF multipliers from empirical data. The paper includes several ratios that can be translated to PSF multipliers, including a linear expression of hours of wakefulness on number of lapses, where lapses increased at a rate of 0.1129 per hour to the baseline of 4.798 ($y = 0.1129x + 4.798$). It should be noted that the linear expression had a fairly low degree of fit ($R^2 = 0.1678$) and that the reviewed studies included in the meta-analysis where performed in a laboratory setting using simple tasks. A large-scale longitudinal study of Swedish workers found that self-reported difficulties in sleeping were a significant predictor of fatal occupational accidents (relative risk = 1.89, 95% CI = 1.22–2.94; Åkerstedt, Fredlund, Gillberg & Jansson, 2002).

3.3.1 Sleep Deprivation as a PSF

None of the HRA methods specifically refer to sleep deprivation. However, it could be to some extent covered by HEART's EPC labeled *disruption of normal work-sleep cycles*

(Williams, 1992, p. 23). The disruption of normal work-sleep cycles is likely to be connected to a reduction in sleep quality and possibly also sleep quantity. The multiplier here is, however, only 1.1 in the original method and 1.2 in the update (Williams & Bell, 2017). This means that if the analyst selects this EPC as one of the three most important EPCs (as the analyst is advised not to choose more than three), and the analyst chooses to give it the full weight, the HEP will be increased by 10% (20% if the updated data are accepted). The *degraded fitness* level of the *Fitness for Duty* PSF in SPAR-H, with a multiplier of five, is possible to use in cases of sleep deprivation as it refers to fatigue and lists examples of causes without limiting the use to only those examples. Fatigue is, however, only a small part of the *Fitness for Duty* PSF, which was originally designed to cover chemically-induced impairment and not fatigue (Griffith & Mahadevan, 2011), and if the PSF is used it is likely to be in a retrospective analysis.

The lack of sleep deprivation in HRA methods is perhaps not surprising. While the effect of sleep deprivation is found in studies, it is difficult to incorporate those findings into a practical tool. Sleep deprivation can be caused by factors at the workplace such as the shift system or very long shifts, but as sleep is generally conducted outside of work hours most causes are outside of the workplace. For example, sleep deprivation can be caused by a sleep disorder, caring for young children, poor planning, or a preference for staying up late. As an HRA is not only a tool for exploring risk, but also a tool in finding how to reduce risk, there is limited value in including an aspect that will vary to such a degree between operators and that the employers can influence to such a small degree. If sleep deprivation were to be included in an HRA method, we believe the only practical way of conducting it would be through a subjective measurement of whether the sleep deprivation has an effect. It is difficult to connect a specific multiplier to a subjective measurement. In our potential Sleep deprivation PSF we have chosen to use the multiplier of five from the method which is the most similar to Petro-HRA, SPAR-H (see Table 1).

Table 1

Levels and Multiplier for a Sleep Deprivation PSF

Degree of sleep deprivation	Multiplier
Sleep deprivation is not performance driving	1
The operator is experiencing tiredness and sleepiness due to a high degree of sleep deprivation	5

3.2 Shift Length

The effect of shift length on HEP was studied in medical interns as two different schedules were compared; the first had the interns work shifts of 24 hours or more every third day, while the other schedule had a reduced work amount with consecutive work periods limited to approximately 16 hours (Landrigan et al., 2004; Lockley et al., 2004). While using the first schedule, interns made 35.9% more serious medical errors and made 5.6 times as many serious diagnostic errors. The degrading effects in human performance in medical interns were seen again in a large-scale survey (2,737 participants provided 17.003 monthly reports). In a related study Barger et al. (2005) studied the effect shift length had on driving performance on the drive home after work. They found that interns were more than twice as likely to be in a car accident (OR = 2.3, 95% CI = 1.6–3.3) and almost six times as likely to experience a near-miss (OR = 5.9, 95% CI = 5.4–6.3) on the drive home after an extended shift (more than 24 hours) compared to a non-extended shift (less than 24 hours) (Barger et al., 2005). A similar negative effect was also found in shorter shift-lengths in a review of three large studies (Nachreiner, Akkermann, & Haenecke, 2000; Haenecke et al., 1998; Åkerstedt, 1995) and a review of smaller studies (Folkard, 1997), were Folkard and Lombardi (2006) found an increased risk of accidents and injuries of 13% and 27.5% in 10- and 12-hour shifts, respectively, when compared to an 8-hour shift. This effect was not seen in shifts ranging from four to nine hours as the relative risk remained relatively stable.

3.3.2 Shift Length as a PSF

Shift length is mentioned (as duty hours) in the *degraded fitness* level of the *Fitness for Duty* PSF in SPAR-H. However, as mentioned before, this PSF is rarely used in prospective analyses. Shift length is an aspect that has extra interest as it is to a large extent within the employer's control. In addition, the results from studies are relatively easy to convert to HEP multipliers and information on the length of a shift will be possible to obtain during the HRA process. The multiplier would, however, be low for the shift lengths that one would realistically come across in the petroleum industry. A shift length PSF (see Table 2) could consist of a nominal level for a shift up to eight hours, a multiplier of 1.13 for shifts lasting 8–10 hours, and a multiplier of 1.28 for shifts lasting 10–12 hours. The multipliers are based on the 13% and 27.5% increase in risk found for 10- and 12-hour shifts compared to an 8-hour shift in Folkard and Lombardi's (2006) review. A PSF that covers shift lengths up till 12-hours would be sufficient for most cases within the petroleum industry. However, it would be an advantage to include longer shifts that could occur under special circumstances. Increased diagnostic errors made by medical interns (Landrigan et al., 2004; Lockley et al., 2004) in shifts lasting more than 24 hours could be used as an input to a 24-hour shift level, unfortunately they used a different baseline (it was compared to 16-hour shifts, the ratio against an 8-hour shift is not known). Another possibility would be to extrapolate the close-to-linear relationship beyond the 12-hour mark (as shown in Table 2). This approach would have weaknesses as there is no guarantee that the close-to-linear relationship would sustain as shift lengths increase. More research on these shift lengths across several domains would be beneficial, although these shift lengths are uncommon in most industries.

Table 2

Levels and Multiplier for a Shift Length PSF

Shift Length	Multiplier
8 hours or less	1
10 hours	1.13
12 hours	1.28
16 hours	1.56
20 hours	1.84
24 hours	1.98

3.3 Non-Day Shift

Three Mile Island, Chernobyl, and Exxon Valdez were mentioned earlier as incidents where sleep deprivation was found to be a partial cause, but they also share another factor; they all occurred at night. This has brought attention to how different shifts, and the night shift in particular, affect human performance and safety. Most people live their lives according to a 24-hour cycle of sleep and wakefulness, and throughout this 24-hour cycle the ratio between sleepiness and wakefulness will vary following an internal clock. This is often referred to as the circadian rhythm. Unfortunately for those working at night, few are lucky enough to have circadian rhythms that easily adapt to frequent changes in the activity schedule or to have a tolerance that allows performance to stay the same when work is conducted outside the hours they are used to (Kuhn, 2001).

There are several challenges in comparing statistics from different shifts as most places of work have different activities at day and at night. Even at platforms or production facilities that operate throughout the day, activities such as maintenance are more likely to occur during the day shift. While maintenance in itself could lead to an accident, maintenance could also mean that there are additional personnel present that could provide the operator with information from the field or assist in the control room depending on the situation. There could also be several variables that influence the personnel on the different shifts. In the U.S., many permanent shift systems operate on seniority, which would often mean that those with the least experience would get the least desirable shift, which is typically the night shift (Folkard & Lombardi, 2006). By the time they move through the afternoon shift and to the day shift they are likely to be the most experienced workers and likely to possess a very different skillset than when they worked nights. Even in places without a permanent shift system and where maintenance activities do not differ between the shifts, other differences can skew the data. In an example mentioned by Folkard and Tucker (2003), there was a large difference between the number of reported non-serious injuries, with more being reported during the day shift. Further investigation showed that injuries during the day were treated by female nurses in an on-site occupational health clinic while injuries at night were treated by male security guards. It seemed probable that this might be the reason for the difference in reported injuries by the predominantly male workforce. The suspicion was enhanced by the fact that the number of injuries also varied substantially depending on which nurse was on duty (Folkard & Tucker, 2003). Other potential cases of such a confounding variable might be seen in the studies of injuries in a Polish steel plant (Oginski, Oginska, Pokorski, Kmita, & Gozdziela, 2000) and Oregon hospitals (Horwitz & McCall, 2004). Both studies found that there were fewer injuries at night (in the case of the Oregon hospitals the injury rate of the night shift was only lower compared to the afternoon shift, while the day shift had the lowest injury rate of the shifts); however, they also both found that the injuries at night were more severe. It could be that the effects of night-shift fatigue increase only the type of risky behavior that leads to serious injuries, or it could be something as simple as that less serious accidents are less frequently reported at night.

In a review Folkard and Tucker (2003) found a close to linear relationship in increased risk from the morning shift to the afternoon shift and then to the night shift. Compared to the morning shift, the afternoon and night shifts had increased risks of 18.3% and 30.4%, respectively. A later review including the same studies but using a slightly different meta-analytical technique found the effect to be 15.2% and 27.9%, respectively (Folkard & Lombardi, 2006). While the reviews only included studies with a relatively large amount of incidents (ranging from 119 to 4,600, with an average of 2,498.2) and the risk increase was present in all studies, it should be noted that while many confounding variables where ruled out in these studies, the review was based on only five studies and that controlling all potential confounding variables is never possible. Similar results were found in the large scale study of Swedish workers where non-daytime work was found to be a significant predictor of accidents (relative risk: 1.63, 95% CI = 1.09–2.45; Åkerstedt et al., 2002).

3.3.3 Non-Day Shift

Time of day (circadian rhythm) is included as a CPC in the CREAM method where the *night (non-adjusted)* level has a multiplier of 1.2 for all types of cognitive functions (Hollnagel, 1998). It was introduced as one of the two additional EPCs in the HEART update: Time-of-day (from diurnal high arousal to diurnal low arousal), with a maximum multiplier of 2.4. Reviews of relevant research indicate that the relationship between different shifts during the day appear to be close to linear, with the afternoon shift having a relative risk of approximately 1.15 and the night shift having a relative risk of approximately 1.3 compared to the day shift (Folkard & Lombardi, 2006; Folkard & Tucker, 2003). The vast majority of the incidents included in the calculation of these multipliers were non-serious incidents, which does lead to some uncertainty to the validity in using these values for tasks conducted in a post-initiator scenario. While the night shift value is not identical to the value found in CREAM, there is not a large discrepancy. The values can appear to be much lower than the HEART multiplier, but as HEART uses a maximum multiplier which is adjusted when used by an analyst, it is difficult to compare the values. As there could be many differences in a scenario if it occurred at night, a non-day shift PSF could be an interesting PSF to include in an HRA method.

Table 3

Levels and Multiplier for a Non-Day Shift PSF

Shift	Multiplier
The task is performed during the day shift	1
The task is performed during the afternoon shift	1.15
The task is performed during the night shift	1.30

3.4 Prolonged Task Performance

Fatigue can be caused by performing a task or remaining highly vigilant for an extended period of time. To avoid the dangers associated with this type of fatigue, many of the industries where a person plays a critical safety role have rules and regulations on the maximum amount of time that can be worked without a break. Such regulations exist within the petroleum industry as well; however, in an accident scenario it would seem plausible that if required these regulations would not necessarily be followed. Most of the research on the effects of prolonged tasks have been conducted in laboratory settings and have been on monotonous tasks (Horne & Reyner, 1999). In one of the few studies using real-life data, Tucker, Folkard and Macdonald (2003) assessed 3 years of accident records from a large engineering company finding that compared to the first half hour of work, the relative risk of an accident occurring increased by 1.33 (95% CI =

1.06–1.60) for the second half hour, by 1.71 (95% CI = 1.40–2.02) for the third half hour, and by 2.08 (95% CI = 1.73–2.43) for the fourth half hour following a rest break, showing a general trend of increased risk as time without a break increases. Break duration has been suggested as a meditator variable in prolonged task performance fatigue with arguments supporting both short and longer breaks (Lim & Kwok, 2016). The breaks in this study ranged from short 10 minute breaks to 45 minute meal breaks. Differences in recuperative effects from the different break lengths were not found in this study.

3.3.4 Prolonged Task Performance as a PSF

Prolonged task performance, or long-lasting events, are to some degree included in HEART in one of the EPCs: *prolonged inactivity or highly repetitious cycling of half hour low mental workload tasks* (Williams, 1992, p. 23) with multipliers of 1.1 for first half hour and 1.05 for each hour thereafter. This EPC could be used to describe prolonged task performance, but since it refers to prolonged inactivity and cycles of low mental workload, it is not well suited to describe the likely situation in a post-initiator situation. A set of values that would be better suited for this situation could be the *minutes since last break* measured through relative risk (Tucker et al., 2003). As the car assembly plant that was assessed had breaks every two hours, the study did not include data beyond two hours. We have chosen to set a conservative value of five for the PSF level *the task is performed for more than 120 minutes without a break* (see Table 4). However, as these values are based on a single study from a single industry, caution should be taken in the validity of these values.

Table 4

Prolonged Task Performance as a PSF

Task length	Multiplier
The task is performed for 0–30 minutes without a break	1
The task is performed for 31–60 minutes without a break	1.33
The task is performed for 61–90 minutes without a break	1.71
The task is performed for 91–120 minutes without a break	2.08
The task is performed for more than 120 minutes without a break	5

4. Discussion

There is a substantial amount of research showing a general consensus that fatigue impairs performance. Moreover, several review studies allow for the results to be converted to PSF multipliers to adjust a nominal HEP. There are, however, several challenges in including one or more fatigue PSFs.

4.1. Fatigue PSF Challenges

The first challenge in including fatigue through PSFs is the choice on whether each fatigue element should be created as a separate PSF, or if all fatigue elements should be combined in one PSF. Both approaches have benefits and drawbacks which are highlighted in the following sections.

4.1.1 Large PSF Taxonomy

If an HRA method includes fatigue through one PSF for each cause of fatigue it could lead to a method with a very high number of PSFs. Not necessarily from the fatigue factors by themselves, but rather from the total number of PSFs included if all aspects are treated this way in the creation of the PSF taxonomy. A high number of PSFs is not necessarily a problem, although if the number becomes high enough that the method is considered to be unwieldy or too time consuming, it increases the chances that it is not applied properly if it is ever used. A possible solution could be to include all fatigue aspects into one PSF; however, that would lead to a complicated situation when more than one fatigue aspect is degrading performance in the same event or task.

4.1.2 Non-Exhaustive Fatigue Taxonomy

In addition to the four aspects of fatigue that are presented in this paper, there are several other factors that can have a great influence on fatigue such as general individual differences in rest needed, how quickly a person experiences the effects of fatigue, drug or alcohol impairment, illness, the social environment at work, noise, food and water intake, personality, age, and personal problems (e.g., economical worries or the loss of a loved one). A possible solution is to include an "other causes of fatigue" aspect to allow an analyst to attribute fatigue which does not fit within the set of aspects. The effect of this collection of aspects on performance would likely have to be assessed by the analyst performing the analysis.

4.1.3 Cumulative Effects

Factors that influence fatigue are unlikely to do so separately. Although a single factor is able to cause fatigue by itself if present in a high enough degree, it is more likely that fatigue is caused by several of these effects simultaneously or over time. Grandjean (1968) used the metaphor of a bucket being filled by all the aspects that contribute to fatigue, and the bucket could only be emptied through rest. The metaphor highlights the problem of not including all aspects that contribute to fatigue. This is also a problem for the studies being conducted on fatigue. Real-world data have the disadvantage that it is difficult to control for all other factors that influence fatigue, while controlled experiments have the disadvantage that they do not add the effects of all the other factors that would contribute to fatigue in a real-world setting.

4.1.4 Mediating and Moderating Effects

For the purpose of suggesting PSFs that are possible to use in an HRA setting, we have suggested PSFs based on the general trend of the fatigue effect on performance. This means that mediating and moderating effects are not taken into account. The lack of knowledge and inclusion of how PSFs interact is however not a problem exclusive to fatigue, but rather a general HRA problem.

4.1.5 Task Differences

Furthermore, not enough is known about which aspects of human performance are affected by fatigue. In this paper, the decrementing effect of fatigue has been considered to generally be equal for different tasks. This is not necessarily the case. For example, several studies have found that fatigue has a particularly decremental effect in tasks where finding an optimal strategy is the goal, and that secondary tasks are compromised to a larger degree than primary tasks (Hockey, 1997; Hockey, Wastell, & Sauer, 1998). Similar results were found in a review where sleep disruption had a higher impact on simple tasks than complex ones (Wickens, Hutchins, Laux, & Sebok, 2015). Further, it seems likely that tasks performed in different settings would be differently affected. It is possible that tasks conducted in a post-initiator scenario would be less affected by fatigue than a monotonous routine task (given that the operator realizes that it is a post-initiator scenario). On the other hand, fatigue could cause a narrow focus, making the operator focus on a small problem that is noticed first, not realizing that something far more important should be done first. More research on how fatigue specifically influences different types of tasks would enable the creation of a more complete picture of how fatigue affects performance. For methods such as Petro-HRA which has a focus on control room tasks, simulator studies including both different types of fatigue as well as different types of tasks could be way to achieve this.

4.2. Recommendation on Including Fatigue

The most accurate inclusion of fatigue through a PSF is likely to be in a way that includes as many fatigue causes, mediators and moderators as possible. This could potentially be done through a complex model within a single PSF; however much more research would be required before such a model would be ready and include suitable values. A simplified inclusion of fatigue where each fatigue cause is a separate PSF (Table 1-4) seems like a more realistic inclusion. However, it should be made clear that this is a simplified inclusion of fatigue with certain limitations.

4.3. Fatigue in Petro-HRA

The first question asked in the creation of a PSF for the PSF taxonomy in Petro-HRA was "Can it have a significant effect on human performance in an accident scenario?". As fatigue is frequently mentioned as a factor in accident investigations it appears to have an effect on performance. The significance of the effect is more difficult to evaluate. The values found in all the studies reviewed in this paper indicate that an appropriate multiplier for one or more fatigue PSFs would be that are lower than those included in the Petro-HRA for other factors, where most included factors can reach a multiplier of at least 10 within a realistic scenario. This combined with the unavailability of many of the important factors that contribute to fatigue led to the decision that fatigue should not be included as a PSF in the Petro-HRA method. The decision is similar to the one made in in THERP (Swain & Guttmann, 1983) where fatigue was not included as a PSF, not because it was found not to affect human performance or risk, but because the effect was not large enough. It is also likely that the effect would be highly influenced by individual differences and to a large degree caused by influences outside of both the scenario and the workplace making it very difficult to capture by an HRA.

5. Conclusion

There is a substantial amount of evidence showing that fatigue will influence performance and it is likely that this would also be the case in post-initiator scenarios in the petroleum industry. However, the effects of fatigue were found not to be strong enough for inclusion in an HRA method which only includes the PSFs that are most likely to be performance drivers.

Acknowledgements

The present study was supported by the PETROMAKS Programme of the Research Council of Norway, Grant no. 220824/E30 “Analysis of human actions as barriers in major accidents in the petroleum industry, applicability of human reliability analysis methods (Petro-HRA)”. Financial and other support from The Research Council of Norway, Statoil ASA and DNV GL are gratefully acknowledged. This paper represents the opinion of the authors, and does not necessarily reflect any position or policy of the above mentioned organizations.

References

Barger, L. K., Cade, B. E., Ayas, N. T., Cronin, J. W., Rosner, B., Speizer, F. E., & Czeisler, C. A. (2005). Extended work shifts and the risk of motor vehicle crashes among interns. *New England Journal of Medicine, 352*, 125–134.

Boring, R. L. (2012). Fifty Years of THERP and Human Reliability Analysis. In *Proceedings of the PSAM 11 and ESREL 2012 Conference on Probabilistic Safety Assessment June 25-29*. Helsinki, Finland.

Boring, R. L. (2015). Defining Human Failure Events for Petroleum Risk Analysis. In *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, Las Vegas, NV* (Vol. 3, pp. 1335–1342). Elsevier B.V. https://doi.org/10.1016/j.promfg.2015.07.289

Boring, R. L., & Blackman, H. S. (2007). The origins of the SPAR-H method's performance shaping factor multipliers. In the proceedings of the *IEEE Conference on Human Factors and Power Plants* (pp. 177–184). doi:10.1109/HFPP.2007.4413202

Bye, A., Laumann, K., Taylor, C., Rasmussen, M., Øie, S., van de Merwe, K., … Gould, K. (2017). *The Petro-HRA Guideline (IFE/HR/F-2017/001)*. Halden, Norway: Institute for Energy Technology.

Folkard, S. (1997). Black times: Temporal determinants of transport safety. *Accident Analysis and Prevention*, *29*, 417–430. doi:10.1016/S0001-4575(97)00021-3

Folkard, S., & Lombardi, D. A. (2006). Modeling the impact of the components of long work hours on injuries and "accidents." *American Journal of Industrial Medicine*, *49*, 953–963. doi:10.1002/ajim.20307.

Folkard, S., & Tucker, P. (2003). Shift work, safety and productivity. *Occupational Medicine*, *53*, 95–101. doi:10.1093/occmed/kqg047

Gertman, D., Blackman, H., Marble, J., Byers, J., & Smith, C. (2005). *The SPAR-H human reliability analysis method.* US Nuclear Regulatory Commission.

Gould, K. S., Ringstad, A. J., & van de Merwe, K. (2012). Human Reliability Analysis in Major Accident Risk Analyses in the Norwegian Petroleum Industry. In *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting, October 22-26, Boston, MA* (pp. 2016–2020).

Graham, B., Reilly, W. K., Beinecke, F., Boesch, D. F., Garcia, T. D., Murray, C. A., & Ulmer, F. (2011). *Deep Water: The Gulf Oil Disaster and the Future of Offshore Drilling. Report to the President. National Commission on the BP Deepwater Horizon Spill and Offshore Drilling*. Washington, DC. https://doi.org/10.3723/ut.30.113

Grandjean, E. (1968). Fatigue: Its physiological and psychological significance. *Ergonomics*, *11*, 427–436.

Griffith, C. D., & Mahadevan, S. (2011). Inclusion of fatigue effects in human reliability analysis. *Reliability Engineering and System Safety*, *96*, 1437–1447. doi:10.1016/j.ress.2011.06.005

Griffith, C. D., & Mahadevan, S. (2015). Human reliability under sleep deprivation : Derivation of performance shaping factor multipliers from empirical data. *Reliability Engineering and System Safety*, *144*, 23–34. https://doi.org/10.1016/j.ress.2015.05.004

Hänecke, K., Tiedemann, S., Nachreiner, F., & Grzech-Šukalo, H. (1998). Accident risk as a function of hour at work and time of day as determined from accident data and exposure models for the German working population. *Scandinavian Journal of Work, Environment and Health*, *24*, 43–48.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*, 73–93.

Hockey, G. R. J., Wastell, D. G., & Sauer, J. (1998). Effects of sleep deprivation and user interface on complex performance: A multilevel analysis of compensatory control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *40*, 233–253. doi:10.1518/001872098779480479

Hollnagel, E. (1998). *Cognitive Reliability and Error Analysis Method CREAM*. Eastbourne, UK: Antony Rowe Ltd.

Horne, J., & Reyner, L. (1999). Vehicle accidents related to sleep: A review. *Occupational and Environmental Medicine*, *56*, 289–294. doi:10.1136/oem.56.5.289

Horwitz, I. B., & McCall, B. P. (2004). The impact of shift work on the risk and severity of injuries for hospital employees: An analysis using Oregon workers' compensation data. *Occupational Medicine*, *54*, 556–563. doi:10.1093/occmed/kqh093

Howitt, D. (2010). *Introduction to qualitative methods in psychology*. Essex, UK: Pearson Education Ltd.

Kirwan, B. (1994). *A guide to practical human reliability assessment*. Boca Raton, FL: CRC Press.

Kuhn, G. (2001). Circadian rhythm, shift work, and emergency medicine. *Annals of Emergency Medicine*, *37*, 88–98. https://doi.org/10.1067/mem.2001.111571

Landrigan, C. P., Rothschild, J. M., Cronin, J. W., Kaushal, R., Burdick, E., Katz, J. T., … Czeisler, C. A. (2004). Effect of reducing interns' work hours on serious medical errors in intensive care units. *The New England Journal of Medicine*, *351*, 1838–1848. doi:10.1056/NEJMoa041406

Laumann, K., & Rasmussen, M. (2016). Suggested improvements to the definitions of Standardized Plant Analysis of Risk-Human Reliability Analysis (SPAR-H) performance shaping factors, their levels and multipliers and the nominal tasks. *Reliability Engineering & System Safety*, *145*, 287–300.

Laumann, K., & Rasmussen, M. (2017). Experience and training as performance-shaping factors in human reliability analysis. In L. Walls, M. Revie, & T. Bedford (Eds.), *Risk, Reliability and Safety: Innovating Theory and Practice. Proceedings of the 26th European Safety and Reliability Conference, ESREL 2016, Glasgow, Scotland, 25-29 September 2016*. CRC Press.

Laumann, K., Øien, K., Taylor, C., Boring, R. L., & Rasmussen, M. (2014). Analysis of human actions as barriers in major accidents in the petroleum industry, applicability of human reliability analysis methods. In the proceedings of the *Probabilistic Safety Assessment and Management (PSAM12),* Honolulu, HI.

Lim, J., & Kwok, K. (2016). The Effects of Varying Break Length on Attention and Time on Task. *Human Factors*, *58*, 472–481. https://doi.org/10.1177/0018720815617395

Lockley, S. W., Cronin, J. W., Evans, E. E., Cade, B. E., Lee, C. J., Landrigan, C. P., … Czeisler, C. A. (2004). Effect of reducing interns' weekly work hours on sleep and attentional failures. *The New England Journal of Medicine*, *351*, 1829–1837. doi:10.1097/01.sa.0000158571.48089.a3

Mitler, M. M., Carskadon, M. A., Czeisler, C. A., Dement, W. C., Dinges, D. F., & Graeber, R. C. (1988). Catastrophes, sleep, and public policy: Consensus report. *Sleep*, *11*, 100–109.

Nachreiner, F., Akkermann, S., & Haenecke, K. (2000). Fatal accident risk as a function of hours into work. In S. Hornberger, P. Knauth, G. Costa, & S. Folkard (Eds.), *Shiftwork in the 21st Century* (pp. 19–24). Frankfurt, Germany: Peter Lang.

Noy, Y. I., Horrey, W. J., Popkin, S. M., Folkard, S., Howarth, H. D., & Courtney, T. K. (2011). Future directions in fatigue and safety research. *Accident: Analysis and Prevention*, *43*, 495–497. doi:10.1016/j.aap.2009.12.017

NUREG-1624. (2000). *Technical Basis and Implementation Guidelines for a Technique for Human Event Analysis (ATHEANA)* (Rev.1). Washington, DC: Division of Risk Analysis and Applications, Office of Nuclear Regulatory Research.

NUREG-1880. (2007). *ATHEANA User's Guide, Final Report.* Washington, DC: Division of Risk Assessment and Special Projects, Office of Nuclear Regulatory Research.

Oginski, A., Oginska, H., Pokorski, J., Kmita, W., Gozdziela, R. (2000). Internal and external factors influencing time-related injury risk in continuous shift work. *International Journal of Occupational Safety Ergonomics*, *6*, 405–421. doi:10.1080/10803548.2000.11076464

Rasmussen, M. (2016). *The Development of Performance Shaping Factors for the PetroHRA Method: A Human Reliability Method for the Petroleum Industry*. Doctoral dissertation, Norwegian University of Science and Technology, Trondheim, Norway.

Rasmussen, M., & Laumann, K. (in review). The evaluation of human-machine interface as a performance shaping factor in the PetroHRA method.

Rasmussen, M., Standal, M. I., & Laumann, K. (2015). Task complexity as a performance shaping factor: A review and recommendations in Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H) adaption. *Safety Science*, *76*, 228–238. https://doi.org/10.1016/j.ssci.2015.03.005

Skogdalen, J. E., & Vinnem, J. E. (2011). Quantitative risk analysis offshore - Human and organizational factors. *Reliability Engineering System Safety*, *96*, 468–479. doi:10.1016/j.ress.2010.12.013

Swain, A. D. (1987). *Accident sequence evaluation program: Human reliability analysis procedure (NUREG/CR-4772).* Albuquerque, NM: Sandia National Labs.

Swain, A. D. (1990). Human reliability analysis: Need, status, trends, and limitations. *Reliability and Systems Safety*, 29, 301–313. doi:10.1016/0951-8320(90)90013-D.

Swain, A., & Guttmann, H. (1983). *Handbook of human-reliability analysis with emphasis on nuclear power plant applications, Final report* (NUREG/CR-1278). Washington, DC: U.S. NRC.

Thorndike, E. (1900). Mental fatigue. *Psychological Review*, *7*, 466–482.

Tucker, P., Folkard, S., & Macdonald, I. (2003). Rest breaks and accident risk. *The Lancet*, *361*, 680.

Van de Merwe, K., Hogenboom, S., Rasmussen, M., Laumann, K., & Gould, K. (2014). Human Reliability Analysis for the Petroleum Industry : Lessons Learned from Applying SPAR-H. *SPE Economics & Management*, 1–7.

Vinnem, J. E. (1998). Evaluation of methodology for QRA in offshore operations. *Reliability Engineering & System Safety*, *61*, 39–52. doi:10.1016/S0951-8320(97)00063-X

Wickens, C. D., Hutchins, S. D., Laux, L., & Sebok, A. (2015). The Impact of Sleep Disruption on Complex Cognitive Tasks : A Meta-Analysis. *Human Factors*, *57*, 930–946. https://doi.org/10.1177/0018720815571935

Williams, J. C. (1988). A data-based method for assessing and reducing human error to improve operational performance. In *Conference Record for 1988 IEEE Fourth Conference on Human Factors and Power Plants*, (pp. 436–450). IEEE. doi:10.1109/HFPP.1988.27540

Williams, J. C. (1992). *A User Manual for the HEART*. Stockport, UK: DNV Technica Ltd.

Williams, J. C., & Bell, J. L. (2017). Consolidation of the human error assessment and reduction technique. In *Risk, Reliability and Safety: Innovating Theory and Practice. Proceedings of the 26th European Safety and Reliability Conference, ESREL 2016, Glasgow, Scotland, 25-29 September 2016* (pp. 883–889). Glasgow, Scotland.

Williamson, A., Lombardi, D. A, Folkard, S., Stutts, J., Courtney, T. K., & Connor, J. L. (2011). The link between fatigue and safety. *Accident: Analysis and Prevention*, *43*, 498–515. doi:10.1016/j.aap.2009.11.011

Øien, K., & Sklet, S. (1999). *Bruk av risikoanalyser i driftsfasen, etablering av sikkerhetsindikatorer og modellering av organisatoriske faktorers effekt på risikonivået - en "state-of-the-art" beskrivelse (in Norwegian)*. Trondheim, Norway: SINTEF.

Åkerstedt, T. (1995). Work injuries and time of day - national data. *Shiftwork International Newsletter*, *12*, 2.

Åkerstedt, T., Fredlund, P., Gillberg, M., & Jansson, B. (2002). A prospective study of fatal occupational accidents–relationship to sleeping difficulties and occupational factors. *Journal of Sleep Research*, *11*, 69–71.