Development and validation of a wearable fatigue monitoring device

Melanie Mertesdorf, Alan Jones & Neil Clark

Integrated Human Factors (IHF ltd), UK

SUMMARY

Fatigue management and in particular associated cognitive depletion is of crucial relevance in areas such as occupational health and transport safety, with some estimating that 25-50% of commercial vehicle accidents occur due to the effects of human error through accumulative cognitive fatigue (Davidović et al., 2018).

The Driver Innovation Safety Challenge (DISC) was commissioned in response to such concerns and recent high-profile incidents. It was a joint project led by Edinburgh Trams with the support of UKTram and Transport for Edinburgh, and a partnership of public and private sector organisations including the City of Edinburgh Council and the Scotland CAN DO Innovation Challenge Fund. Its remit was to promote the development of technology and processes to help mitigate against the onset of mental fatigue and aid in the prevention of associated incidents. Key criteria for this type of management were to enable real time fatigue monitoring of personnel in an unobtrusive manner.

This paper discusses the development and validation of a wrist-worn wearable device for fatigue detection and alerting, with a particular emphasis on validation studies conducted with tram operators in an operationally representative simulator environment. Despite initial focus on tram operators, the device has potential applications for many other domains where mental fatigue could lead to catastrophic events.

KEYWORDS

Fatigue Management, Safety, Human Factors

Introduction

Definitions of Fatigue

Fatigue is often characterized as a reduction in mental and/or physical performance, resulting from cognitive overload (extended physical and/or mental workload), underload (boredom), physical exertion, sleep deprivation, disruption of circadian rhythm, or illness (Mohanavelu et al., 2017). Although numerous definitions of fatigue have been proposed, the multidimensional nature of fatigue, the interaction of various variables, and in many instances, the subjective perception of fatigue, have precluded a unified definition.

Fatigue as an Occupational Hazard

Fatigue impairs cognitive and/or motor performance, reduces situational awareness, work efficiency, productivity, and increases the likelihood of human error, thereby raising the risk of injury and fatality (Yung, 2016). This physical and mental impediment is of utmost significance for workplace health and safety in sectors such as transportation, heavy industry, aviation, and construction. Fatigue affects individuals differently and can cause a range of symptoms, potentially causing individuals to misjudge their fatigue level and the associated risks. Extended working hours, and the resultant fatigue and stress, are associated with an increase in the injury hazard rate among workers (Dembe, 2005).

Chinese police records indicate that 1–4% of road crashes occur due to sleepiness/fatigue (Li et al., 2018). However, this understates the impacts of fatigue, partially due to an inability to assess fatigue at a crash scene. Questionnaires, observations, and in-depth investigation indicate that the actual value is 10–20% (European Commission, 2018), rising to 20–50% when considering only commercial vehicle accidents (Davidović et al., 2018). In the UK, fatigue has been implicated in 20% of accidents on major roads, costing the £115 - £240 million per year in terms of work accidents (HSE). Moreover, fatigue is implicated in 4–8% of aviation incidents (Caldwell, 2005), and a meta-analysis of 27 observational studies estimated up to 13% of workplace injuries could be attributed to fatigue (National Safety Council, 2014).

Driver Innovation Safety Challenge

Following a fatal tram incident in 2016, in which fatigue was identified as a contributory factor (RAIB, 2020), the Driver Innovation Safety Challenge (DISC) was launched to commission a non-intrusive safety device capable of detecting when a tram driver might be about to lose consciousness or focus due to illness or fatigue.

Wearable Devices as a Solution

Given the significant role fatigue plays in accidents across various sectors, it is evident that there is still room for improvement in current fatigue management processes. In particular, additional methods for the detection and management of fatigue onset are required.

Wearable technologies promise both continuous monitoring and real-time feedback of fatigue, which could complement and enhance current proactive fatigue management strategies. They offer the possibility of both reactive measures in response to detected fatigue and improved proactive interventions such as altered scheduling based on observed fatigue patterns (e.g. to target fatigue hotspots).

Wearable devices are already increasingly being incorporated into the workplace for other purposes, to enhance worker effectiveness and productivity, to monitor wellbeing, and to help to ensure safety. As technologies continue to evolve, it is expected that the use of wearables in occupational settings will continue to expand (Deloitte Insights, Schatsky & Kumar, 2021).

Design and Development of the "Baseline NC" Wearable Device

The wearable device (referred to as FOCUS+ for the DISC project) is being marketed as Baseline NC and is made up of key hardware components such as wireless connectivity modules, sensors, accelerometers, and gyroscope. Figure 1 describes the overall system design of the wearable device and hub in an occupational context.

With its fatigue prediction algorithm, this device is designed for continuous estimation of an individual's fatigue level in their occupational setting as part of a robust fatigue management strategy. The system was developed and tested with the assistance of light rail staff, and verified by an independent assessment commissioned by the Light Rail Safety and Standards Board (LRSSB).

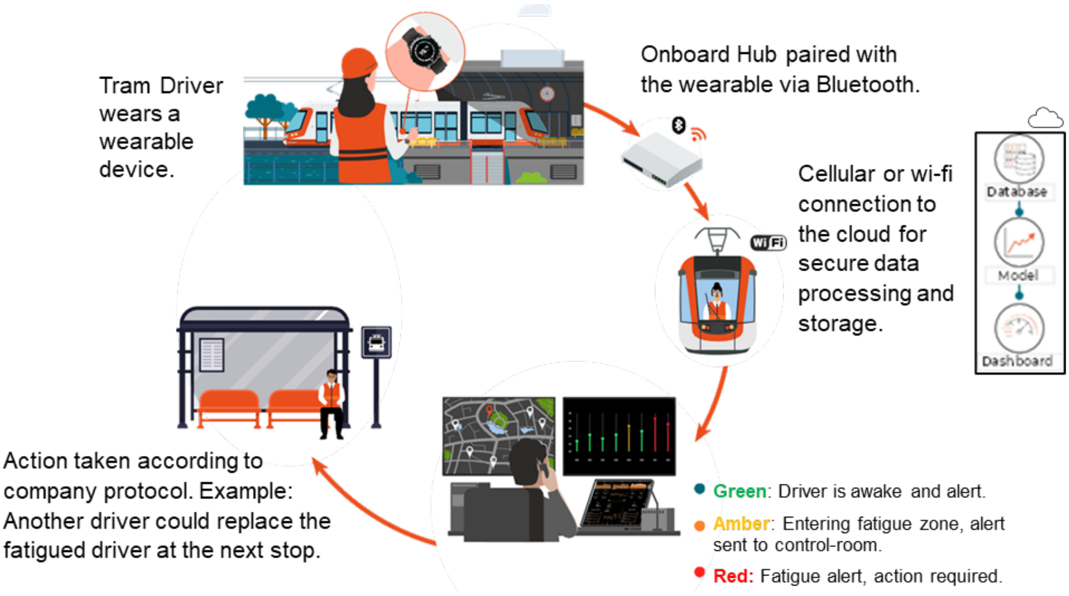

Figure 1: Final system design of the Baseline NC wearable device, using the example of deployment in a tram setting. Data is collected from the driver by the wearable, then transmitted to the hub and on to the cloud. Data is stored and processed in the cloud, with relevant alerts displayed to control room staff or drivers, allowing action to be taken in line with the fatigue management processes (e.g. making contact with the driver, increased monitoring, or intervention and driver handover).

Development and validation of a Fatigue Prediction Algorithm

Initial system development involved the deployment of prototype devices with tram drivers during their normal duties. This provided valuable insights into the practicalities of using the device in an occupational setting. Development of the fatigue prediction algorithm used data from volunteers in tram simulators, which allowed for acquisition of fatigue data from trained staff in a realistic and task-specific setting.

The aim was to develop an actionable predicted RAG fatigue status (Red: fatigue threshold; Amber: approaching threshold; Green: normal) for tram drivers from wearable data.

Fatigue is a multi-modal phenomenon that cannot be narrowly defined. The methods used to develop and validate Baseline NC fatigue estimation algorithm broadly fall into the following key categories.

1. Subjective Measures: This includes self-report questionnaire and scales, such as the Karolinska Sleepiness Scale (KSS), it is an estimate of situational sleepiness, which is an industry-accepted approach to estimating fatigue. Originally proposed by Akerstedt and Gillberg (1990),it is a subjective measure that assesses self-reported fatigue on a 9-point Likert scale and is particularly adept for assessing fatigue relating to shift work, jetlag, and driving abilities, and has been validated by electroencephalogram (EEG) studies (Adão Martins et al., 2021).

2. Performance-Related Methods: These involve assessing the impact of fatigue on physical and cognitive performance. Examples include reaction speed or timed decision-making tests such as the Psychomotor Vigilance Task (PVT). Methods of measuring physical/muscular fatigue (e.g. jump tests, maximal and submaximal sprints, and postural sway measurements) also fall into this category (Hughes et al., 2019). It is an objective measure of sustained attention that relies on the premise that a subject’s cognitive and,

consequently, motor performance on specific tasks is reflective of their fatigue level. It requires attention, and characterizes a person's alertness, attentiveness, and preparedness to receive important information that has not yet arrived. Acceptable levels of vigilance are required for many safety-critical activities including operating a tram. The version of the PVT used here was developed and published by Reifman et al., 2018.

3. Behavioural-Based Methods: These adopt an observational approach to detect fatigue and include external signs, such as yawning, sighing, head nodding, and eye closure – the latter particularly relating to microsleep. Microsleep refers to very brief periods of sleep that can be measured in seconds, rather than minutes or hours, when a person's brain is not processing external information as usual and most likely to occur after sleep deprivation. As a result, people show a reduced response to external stimuli, such as sound or visual cues; they also display less accurate physical interactions.
4. Physiological Signal-Based Methods: These methods measure physiological signals such as heart rate parameters (as performed by the Baseline NC device).

The developed wearable system is an example of a physiological signal-based method. The development and validation processes employed a combination of subjective methods (primarily KSS), performance-related methods (PVT), and behavioural-based observations. Sleep-wake history was also recorded. This combination of methods provided evidence of the validity of fatigue estimates, and allowed for a more holistic assessment of subject fatigue.

The validation study utilised a high-fidelity tram simulator operated by professional tram drivers wearing the Baseline NC throughout. Drivers conducted realistic tram routes and provided KSS scores at defined points, before, during and after each session. Drivers also performed the PVT following each session, and behavioural observations were conducted throughout. Each driver completed a total of 12 hours simulator driving across 4 sessions.

Discussion

Comparison of KSS and PVT results

Predictably, the results of the validation study showed a marked decrease in objective/performance-based measures (PVT) at higher KSS levels, with PVT reaction times during slowest for KSS ratings of 8 and 9, followed by KSS 6 and 7 and quickest for KSS ratings <= 5 (Figure 2, left).. While KSS and PVT results are complementary rather than interchangeable, and there is no conversion method for results, there has been recent work on biomathematical models to allow both KSS and PVT results to be predicted from sleep-wake history (McCauley et al., 2021).

Similarly, Ahrens et al. (2022) observed a "moderate but highly robust" association between KSS score and PVT reaction time over the 1-year time period of their study. These results provide confidence to our use of KSS as a routine, frequently-collected, and reliable fatigue measure for algorithm refinement over time.

Figure 2 (right) also shows an increased number of minor lapses (reaction times >= 500 ms) at KSS 8&9. This is less striking but perhaps more important in an occupational setting, as instances of overly slow reaction times by individuals operating vehicles can easily lead to injury or death of the operator, passengers, or bystanders.

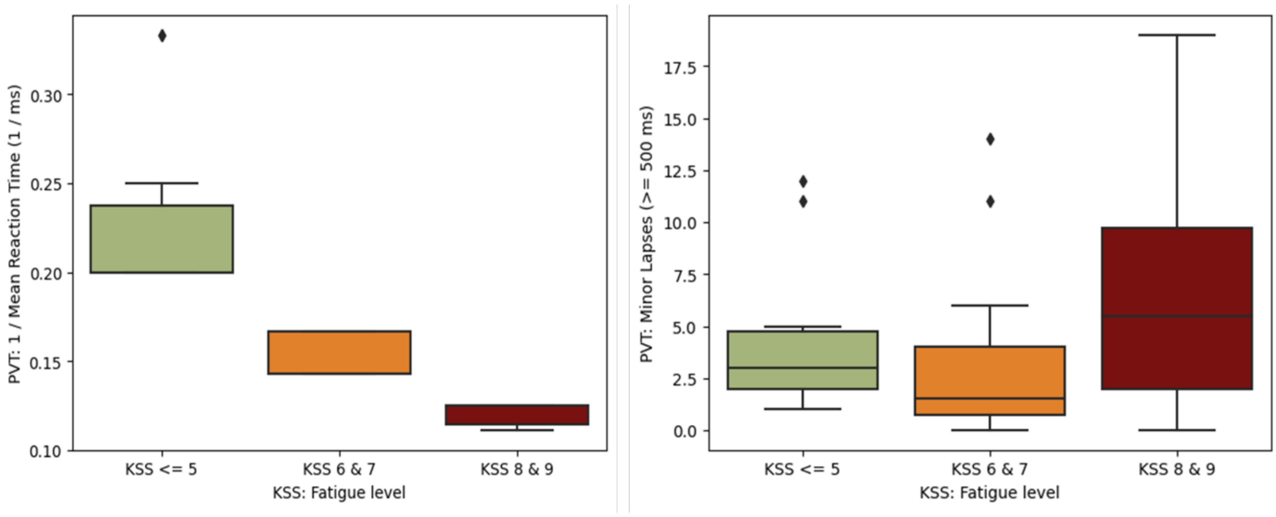

Figure 2: Comparison of KSS and PVT results of tram drivers. KSS results (x-axis) are grouped into three levels reflecting selected red-amber-green alert thresholds. PVT results (y-axis) show the inverse of the mean reaction times in milliseconds (often referred to as 1/RT) on the left graph, and the number of minor lapses per 10-minute PVT session (defined as a reaction time equal to or greater than 500 milliseconds) on the right graph.

Comparison of KSS results and Behavioural Observations

Anderson et al. (2023) report that drivers are typically aware of sleepiness, and recommend self-assessment of a wide range of sleepiness symptoms. Notably however, during a tram simulator session monitored by a trained observer, one subject was seen to exhibit two episodes of microsleep despite self-assessing as "safe to drive" at a self-reported KSS of 8. Task-related decreases in performance were also observed in this period (overshooting a tram stop in the simulated route). A "not safe to drive" declaration was not made by the driver until 5 minutes after the second observed microsleep (Figure 3). This highlights the potential fallibility of even experienced professionals such as tram drivers in estimating their own fatigue during routine tasks.

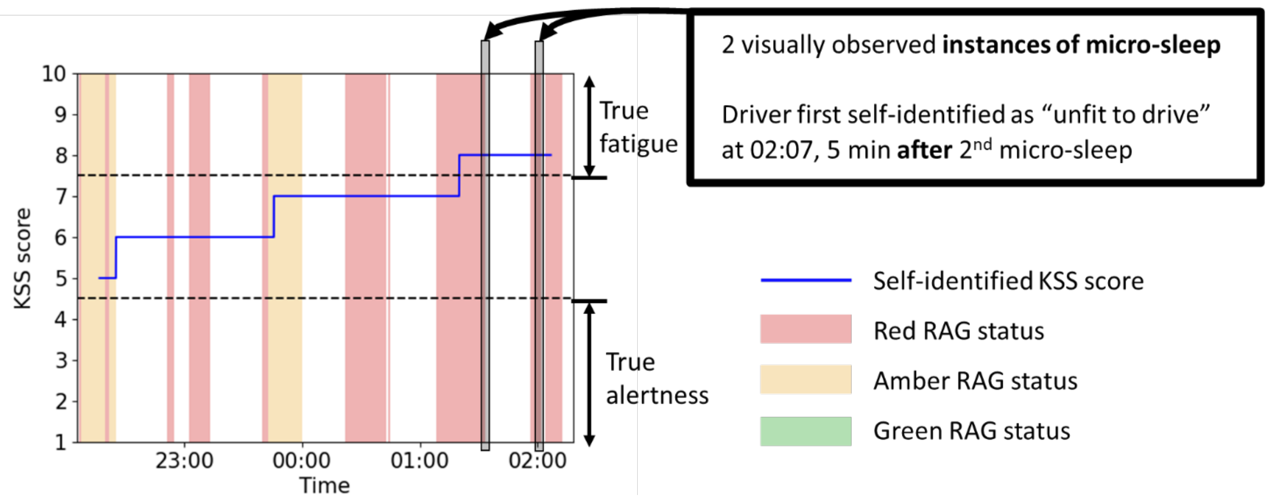

Figure 3: Evolution of KSS score (blue line) and "RAG status" (wearable alert; shaded background colour) of a participant during a night-time session with observed micro-sleep occurrences. White background shading indicates periods of PVT testing and wellness breaks during which the participant was asked not to wear the device.

Establishing the boundaries of fatigue for a fatigue alert system

While KSS ratings formed the major subjective fatigue measure in this study, tram drivers were also intermittently asked whether they considered themselves "safe to drive a tram". There was variation in the subjective onset of fatigue; the distribution (Figure 4) suggests that drivers always considered themselves dangerously tired at a KSS of 9, and often at KSS 8, but only sometimes at KSS 7.

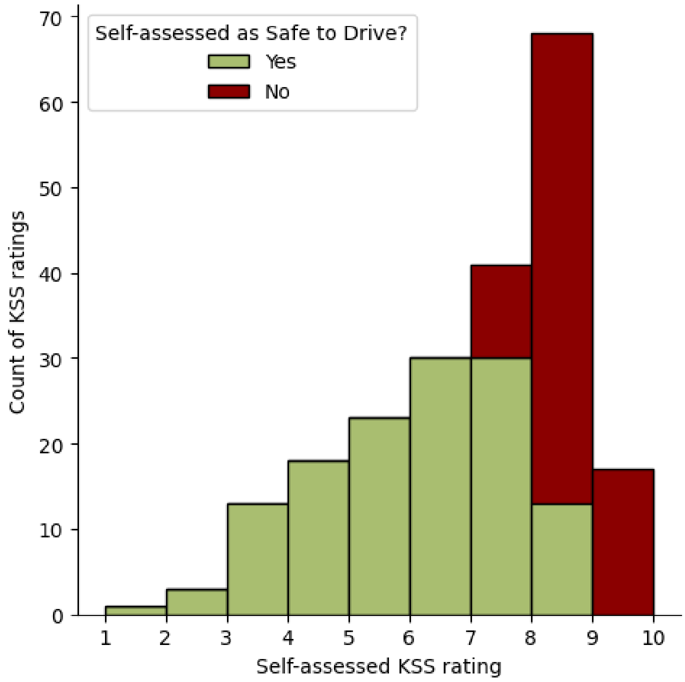

Figure 4: Bar chart showing the prevalence (y-axis) of positive and negative answers by tram driver subjects when asked if they consider themselves "safe to drive a tram", organised by their KSS score at the time (x-axis).

This is in agreement with the accompanying PVT testing (Figure 2), which showed a generally decreased performance at KSS 8 and 9. Ingre et al. (2006) found that crashes and incidents in a driving simulator almost always occurred at KSS 8-9, and Anund et al. (2008) recorded a mean KSS of 8.1 immediately before rumble strip hits in a different driving simulator study. Based on literature results and the evidence given by tram drivers in this study, it was determined that the wearable algorithm should provide a "red alert" for fatigue states corresponding to KSS 8 and 9.

Given the decreased PVT performance at KSS 6 and 7, and the onset of "not safe to drive" judgements at KSS 7, it was decided that an estimated physiological state corresponding approximately to KSS 5-7 should be the threshold for an "amber" fatigue alert by the wearable algorithm, signifying a possible transition into fatigue. Early warnings of possible fatigue could be used to prompt a more detailed self-assessment.

These alert boundaries also agree with guidance from Network Rail, who use the KSS as part of their fatigue reduction strategy. They suggest that anyone scoring 6 or 7 on the KSS needs to self-monitor for symptoms of fatigue, nap if possible, and consider "strategic use of caffeine". Regarding individuals self-assessing as 8 or 9 on the scale, they state that their alertness has reduced to a level where they should not perform safety critical duties including driving.

Conclusion

The discussed results were gathered in a realistic simulator setting, using a novel wrist-worn wearable that allows continuous fatigue prediction. They illustrate the development process of a fatigue management technology, and demonstrate its potential application as part of a broader fatigue management system.

The combination of the comprehensive IHF and independent assessment results show a significant correlation and accuracy (87.5-98%, according to the independent testing) between real life individual fatigue levels, including associated cognitive and physical characteristics, and those indicated by the Baseline NC^{TM} algorithm RAG status.

Thes validation studies provide operational evidence that the system delivers effective monitoring of fatigue onset. It shows the device can be used as an effective tool to minimise potential catastrophic events associated with human error, linked to fatigue. Its applications are numerous within the workplace environment, with functionality having the capability to be adapted to meet the operational domains of various industries.

References

Adão Martins, N. R., Annaheim, S., Spengler, C. M., & Rossi, R. M. (2021). Fatigue monitoring through wearables: A state-of-the-art review. Frontiers in physiology, 12, 2285.

Ahrens, E., Hemmsen, M. C., Duun-Henriksen, J., Kjær, T. W., Allen, L., Cormack, F., & Taptiklis, N. (2022). 0020 Investigation of Seasonal Changes in Self-Reported Sleep Quality and Psychomotor Vigilance Task Outcomes: Results from the Ultra Long-Term Sleep (ULTS) Study. Sleep, 45(Supplement_1), A8–A9

Akerstedt T, Gillberg M. (1990). Subjective and objective sleepiness in the active individual. International Journal of Neuroscience, 52, 29–37.

Anderson, C., Cai, A. W., Lee, M. L., Horrey, W. J., Liang, Y., O'Brien, C. S., ... & Howard, M. E. (2023). Feeling sleepy? stop driving—awareness of fall asleep crashes. *Sleep*, zsad136.

Anund, A., Kecklund, G., Vadeby, A., Hjälmdahl, M., & Åkerstedt, T. (2008). The alerting effect of hitting a rumble strip—A simulator study with sleepy drivers. *Accident Analysis & Prevention*, *40*(6), 1970-1976.

Caldwell J. A. (2005). Fatigue in aviation. Travel Med. Infect. Dis. 3, 85–96. 10.1016/j.tmaid.2004.07.008.

Colston, H. L., & Katz, A. N. (2004). Figurative language comprehension: Social and cultural influences. Mahwah, NJ: Erlbaum.

Davidović J., Pešić D., Antić B. (2018). Professional drivers' fatigue as a problem of the modern era. *Transp. Res. F Traffic Psychol. Behav.* 55, 199–209. 10.1016/j.trf.2018.03.010.

Dembe A. E. (2005). The impact of overtime and long work hours on occupational injuries and illnesses: new evidence from the United States. Occup. Environ. Med. 62, 588–597. 10.1136/oem.2004.016667.

Deloitte Insights, 2021: https://www2.deloitte.com/us/en/insights/focus/signals-for-strategists/wearable-devices-in-the-workplace.html

European Commission (2018). *Fatigue*. Available online at: https://ec.europa.eu/transport/road_safety/sites/default/files/pdf/ersosynthesis2018-fatigue.pdf (accessed March 30, 2021).

HSE. https://www.hse.gov.uk/humanfactors/topics/fatigue.htm

Hughes, S., Chapman, D. W., Haff, G. G., & Nimphius, S. (2019). The use of a functional test battery as a non-invasive method of fatigue assessment. PloS one, 14(2), e0212870.

Ingre, M., Åkerstedt, T., Peters, B., Anund, A., Kecklund, G., & Pickles, A. (2006). Subjective sleepiness and accident risk avoiding the ecological fallacy. *Journal of sleep research*, *15*(2), 142-148.

Li Y., Yamamoto T., Zhang G. (2018). Understanding factors associated with misclassification of fatigue-related accidents in police record. *J. Safety Res.* 64, 155–162. 10.1016/j.jsr.2017.12.002.

McCauley, M. E., McCauley, P., Riedy, S. M., Banks, S., Ecker, A. J., Kalachev, L. V., & Van Dongen, H. P. (2021). Fatigue risk management based on self-reported fatigue: Expanding a biomathematical model of fatigue-related performance deficits to also predict subjective sleepiness. *Transportation research part F: traffic psychology and behaviour*, *79*, 94-106.

Mohanavelu K., Lamshe R., Poonguzhali S., Adalarasu K., Jagannath M. (2017). Assessment of human fatigue during physical performance using physiological signals: a review. Biomed. Pharmacol. J. 10, 1887–1896. 10.13005/bpj/1308.

Network Rail. Fatigue Reduction: How Alert Are You?. https://safety.networkrail.co.uk/wp-content/uploads/2021/11/Fatigue-Fact-Sheet-How-alert-are-you.pdf

NICE Clinikal Knowledge Summaries, 2021: https://cks.nice.org.uk/topics/tiredness-fatigue-in-adults/diagnosis/investigations/

National Safety Council. (2014). Fatigue in the Workplace: Causes and Consequences of Employee Fatigue

RAIB (2020). Rail Accident Report, Overturning of a tram at Sandilands junction, Croydon, 9 November 2016, RAIB Report 18/2017 v 2.2, October 2020

Reifman, J., Kumar, K., Khitrov, M. Y., Liu, J., & Ramakrishnan, S. (2018). PC-PVT 2.0: An updated platform for psychomotor vigilance task testing, analysis, prediction, and visualization. Journal of neuroscience methods, 304, 39-45.

Statista (2017). B.I. Wearable Device Sales Revenue Worldwide from 2016 to 2022 (in Billion U.S. Dollars); Statista Inc.: New York, NY, USA, 2017.

Yung M. (2016). Fatigue at the Workplace*: Measurement and Temporal Development*. Thesis, UWSpace.

Yussuff, V., Sanderson, R. (2014). The World Market for Wireless Charging in Wearable Technology; IHS: Englewood, CO, USA.