How tired is too tired? A study of sleepiness and fatigue incidents reported among UK airline pilots and implications for policy and practice

By Claire Watt-Coombes

A Doctoral Thesis
Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of University College London (UCL)

Declaration

I, Claire Isobel Watt-Coombes confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The risk of fatigue arising from homeostatic (sleep) and circadian (body-clock) drives remains an intrinsic threat to human performance. In ultra-safe high risk (USHRI) industries such as commercial aviation, the question of ‘how tired is too tired for safe flight?’ is one of systemic practitioner importance, requiring robust empirical evidence and conceptual clarity in the way pilot fatigue risks are assessed. This PhD contributes to the field of applied fatigue and safety research in aviation by seeking to understand:

1) What fatigue risk insights can be drawn from mandatory occurrence report (MOR) data when fatigue-related safety occurrences and serious incidents occur?
2) What are the levels of predicted fatigue risk exposure and reported in-flight sleepiness associated with routine UK airline operations?
3) How do fatigue risk exposure and severity findings compare against established sleep science and aviation safety standards for the human safety component?

These questions have been addressed through two studies. Study 1 analysed the quality and coverage of fatigue-related safety occurrence data captured by the main UK safety incident database. Investigation revealed that this system does not capture sufficient data to enable quantitative analysis of the relationship between fatigue risk parameters and safety outcomes. However, through qualitative analysis a convergence of sleep and roster-related themes was evident. Study 2 investigated predicted fatigue risks and reported fatigue experience using bio-mathematical modelling and pilot sleepiness ratings. These methods indicated elevated fatigue risk exposure at both the schedule and individual level. Pilots also reported involuntary sleep on the flight deck at a rate which greatly exceeded the maximum acceptable rate for medical incapacitation in commercial aviation and the rate reported to the regulator. The implications of these research methods and findings are discussed against the context of applied sleep science and aviation safety standards for the human component.

Impact statement

Operator fatigue represents a leading safety concern in aviation. However, current practice is flawed because of a lack of conceptual clarity on what should be reported as well as suggestive evidence of widespread under-reporting. This underscores the need for agreement on a set of meaningful reporting measures and acceptable limits of fatigue risk exposure that are informed by scientific evidence and real world data. Through two studies, this thesis sets out the case for utilising the theoretical construct of the fatigue risk trajectory as a conceptual framework for investigating both distal and proximal fatigue risk exposure, and for developing a common core of key fatigue metrics that can be adopted across the UK commercial aviation sector.

Study 1 shows how the framework can be used to offer new insights on the limitations of the UK MOR safety occurrence database and its effectiveness in identifying fatigue risk data at different organisational defence levels. The study provides the methodological rationale for quality improvement in safety occurrence reporting via structured guidance and mandated data fields. It proposes a suggested set of risk parameters for the Civil Aviation Authority to consider taking forward. The methodological contribution of Study 2 has been to demonstrate the power of combining bio-mathematical modelling of predicted roster-related sleepiness with the generation of subjective sleepiness data from pilots, facilitated by the development for this study of an App that enabled large scale time-stamped data capture. The fact that App technology could enable extensive voluntary research engagement from a remote workforce has important broader implications for future research on fatigue, and the high pilot uptake has been discussed with other industry researchers and bio-mathematical model vendors in aviation. This App and associated study have been highly commended by the Trades Union Congress. I have also presented the findings of Study 2 to several different forums including the Royal Aeronautical Society Fatigue Science Panel.

Many safety practitioners face the challenge of interpreting the severity of fatigue risks in the context of ultra-safe industries such as commercial aviation. The findings of Study 2, as published in *Safety Science,* were the first known to provide per flight hour predicted and reported rates of sleepiness and sleep events across UK airline pilots. The significance of this per flight hour denominator is that it allows new fatigue risk data to be compared against existing safety thresholds for pilot functioning such as the target medical incapacitation rate used by aviation medical doctors. The findings provided by Study 2 were presented at the Global Aviation Health Conference (2019) run by the International Airline Medical Association. This research has generated interdisciplinary interest and to date has been cited by nine papers in medical, safety and transportation fields.

In addition to the conceptual and methodological contributions noted above, the practical, and internationally applicable contribution of this thesis to airline operations management, regulators, pilots, and aviation medical doctors is that it creates the potential for a continuous improvement feedback loop between pilot experience, safety occurrence reporting, problem identification and roster redesign.

Research Paper Declaration

This form is to declare that parts of Study 2 in this thesis are published in a peer-reviewed publication.

Paper reference:

Coombes, C., Whale, A., Hunter, R & Christie, N. (2020). Sleepiness on the flight deck: Reported rates of occurrence and predicted fatigue risk exposure associated with UK airline pilot work schedules. *Safety Science*, 129, Article 104833 https://doi.org/10.1016/j.ssci.2020.104833

The pre-print version of the paper may also be found as an author accepted manuscript at UCL Discovery. https://discovery.ucl.ac.uk/id/eprint/10098928/

1. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
a) Where was the work published? (e.g. journal name)	Safety Science	
b) Who published the work? (e.g. Elsevier/Oxford University Press):	Elsevier	
c) When was the work published?	2020	
d) Was the work subject to academic peer review?	Yes	
e) Have you retained the copyright for the work?	Yes	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
a) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
b) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	

c) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.
d) Stage of publication	Please select.

3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**
This paper forms part of the research for Claire Coombes' PhD, for which she is enrolled as a part-time research student at University College London. This was communicated to the Editor for Safety Science on submission and is noted in the published paper. Claire was solely responsible for the whole-draft write up, analysis, editing and submission of this paper. Editing amendments and suggestions for the paper were provided by co authors. Professor Nicola Christie is Claire's academic supervisor at UCL. Dr Rob Hunter is Claire's industrial supervisor from BALPA. Dr Whale during his time employed at BALPA assisted Claire in the manual input of rosters and provided research tuition to Claire during the study period.
4. In which chapter(s) of your thesis can this material be found?
Chapters 1 and 4
5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	*Claire Watt-Coombes*	**Date:**	22/09/2022
Supervisor/ Senior Author (where appropriate)**:**	*Nicola Christie*	**Date:**	22/09/2022

Acknowledgements

First and foremost, I would like to express immense gratitude and thanks to Professor Nicola Christie for her academic supervision, wise counsel, continued support and encouragement throughout the duration of this work. I would also like to thank Dr Rob Hunter, for igniting my interest in pilot fatigue and providing me with invaluable medical, regulatory and aviation practitioner guidance in his role as industrial supervisor.

I would like to give a special thank you to Dr Andy Whale, who not only provided assistance in my data collection, but also offered unwavering patience, enthusiasm and humour in his sleep science and statistical tutelage, which has been a hugely motivating force for me throughout this PhD.

I would like to express my gratitude to the British airline pilots who participated in the research and BALPA for their support throughout the study period.

Finally, I would like to thank my family for their encouragement and support over the years and my husband, for always believing in me.

Table of Contents

***Declaration* 2**

***Abstract* 3**

***Impact statement* 4**

***Research Paper Declaration* 6**

***Acknowledgements* 8**

***Table of Contents* 9**

***Abbreviations* 11**

***List of Figures* 13**

***List of Tables* 15**

***Chapter 1. Literature Review* 16**

1.1 Introduction 17
1.2 Sleep science Background 21
1.3 Shift factors influencing sleepiness 29
1.4 Consequences of sleepiness 40
1.6. Sleep deprivation and pilot performance 53
1.7 Safety consequences of pilot fatigue 58
1.8 Safety standards and the human component within aviation 62
1.9 Incidence of Fatigue and Sleep-related phenomena in commercial aviation 75
1.10 Sleep Science and recent developments in Safety Theory 77
Conclusion 82

***Chapter 2. Methods* 84**

2.1 The theoretical approach 85
2.2 Aims and objectives 91
2.3 Measurement of sleepiness 94
2.4 Biomathematical modelling of fatigue risk 105

***Chapter 3. Coverage of fatigue risk data in Mandatory Occurrence Reports* 110**

3.1 Ethics Statement 112

3.2 Study Procedure 112
3.3 Results 119
3.4 Discussion of Study 1 Results 131
3.5 Strengths and Limitations 140
Chapter 4. Reported and predicted fatigue risk exposure of Pilot work hours 144
4.1 Ethics Statement 145
4.2 Development of the research protocol 147
4.3 Stakeholder Consultation over App development 151
4.4 The BALPA 2-Way App 159
4.5 Study Procedure 163
4.6 Results 166
4.7 Discussion of Study 2 results 180
4.8 Strengths and limitations 187
Chapter 5. General Discussion 193
5.1 Research approach 194
5.2 Evaluation of study objectives 196
5.3 Implications for policy and practice 200
5.4 Further research 215
Conclusion 219
References 221
Appendices 264
Appendix i. Description of a pilot’s duty 265
Appendix ii. Practical Implications for reporting requirements 267
Appendix iii. Main R packages for analysis 268

Abbreviations

ATC	Air Traffic Control
AME	Aviation Medical Examiner
ATSB	Australian Transport Safety Bureau
BALPA	British Airline Pilots' Association
BST	British Summer Time
CAA	Civil Aviation Authority (UK)
CAS	Circadian Alertness Simulator
CASA	Civil Aviation Authority (Australia)
EASA	European Aviation Safety Agency
EEG	Electroencephalography
EOG	Electrooculography
ESS	Epworth Sleepiness Scale
FAA	Federal Aviation Administration
FAID	Fatigue Assessment Tool by InterDynamics
fMRI	Functional Magnetic Resonance Imaging
FRMS	Fatigue Risk Management System
FRT	Fatigue Risk Trajectory
FTL	Flight Time Limitations
IATA	International Air Transport Association
ICAO	International Civil Aviation Organisation
IR	Infrared
KSS	Karolinska Sleepiness Scale
LED	Light emitting diode

LH	Long-haul
MOR	Mandatory occurrence safety report
NASA	National Aeronautics and Space Administration
NTSB	National Transportation Safety Board (USA)
NZCAA	Civil Aviation Authority (New Zealand)
OSA	Obstructive Sleep Apnoea
PERCLOS	Percentage of eyelid closure over the pupil over time
PVT	Psychomotor Vigilance Task
SH	Short-haul
SAFE	System for Aircrew Fatigue Evaluation
SAFTE	Sleep, Activity, Fatigue and Task Effectiveness (model)
SCM	Swiss Cheese model
SMS	Safety Management System
SP	Samn-Perelli Scale
SSS	Stanford Sleepiness Scale
SWP	Sleep-Wake Predictor Model
SWS	Slow Wave Sleep
SWAI	Sleep Wake Activity Inventory
UTC	Coordinated Universal Time
TSD	Total Sleep Deprivation
TUC	Trades Union Congress
VAS	Visual Analogue Scale
WOCL	Window of Circadian Low

List of Figures

Figure 1	Conceptual clarification of sleepiness and fatigue terms from academic and applied field literatures	20
Figure 2	Summary of the medical incapacitation standard derivation	64
Figure 3a	A basic illustration of Reason’s (1990, 2000) Swiss Cheese Model of accident opportunity	85
Figure 3b	The Swiss Cheese Model	86
Figure 4	An illustration of Dawson & McCulloch’s (2005) Fatigue-risk trajectory (FRT).	87
Figure 5	Summary of the research approach	93
Figure 6	Research Process stages undertaken in Study 1	113
Figure 7	Fatigue-related MORs submitted to the UK Civil Aviation Authority 2014-2019	119
Figure 8	Proportion of MORs containing fatigue risk data at FRT levels 1-3	121
Figure 9	Fatigue symptoms mentioned in MORs	126
Figure 10	Research process stages undertaken in Study 2	146
Figure 11	Data collection opportunities for different sleepiness or operator fatigue measures during a pilot work shift	149
Figure 12a	The BALPA 2-Way App	159
Figure 12b	Screenshots from the BALPA 2-way app	160
Figure 12c	Screenshots from the BALPA 2-way app	161
Figure 12d	Screenshot from the BALPA 2-way app	162
Figure 13	Flow chart indicating participant numbers for different categories of data	166
Figure 14a	Overall frequency of KSS ratings during flying and non-flying hours for long and short haul pilots.	172
Figure 14b	Frequency of KSS ratings plotted in six 4-hour bins across the 24 hour period. Time of day is expressed in British Summer Time (BST).	172
Figure 15	Box plot comparing interquartile range of average KSS ratings by total ratings submitted	173

Figure 16a	Frequency of in-flight ratings by reported KSS	174
Figure 16b	In-flight KSS ratings expressed as a proportion of all submitted ratings by pilot type	174
Figure 17	Predicted KSS and Reported KSS scatterplot	177
Figure 18	Study 2 headline findings of predicted and reported fatigue risk exposure rates	199
Figure 19	Data insights required by FRMS relating to the safety impact of fatigue	211
Figure 20	Appendix i: Illustration of a pilot's duty	265

List of Tables

Table 1	Summary of Neurobehavioural and Cognitive Performance Effects of Sleep Deprivation	45
Table 2	Comparison of Safety I and II theoretical propositions against sleep science methods applicable to commercial aviation practice	80
Table 3	The Karolinska Sleepiness scale	103
Table 4	Coding structure for the assessment of coverage in report narratives	116
Table 5	Fatigue Risk Trajectory framework and corresponding fatigue risk metrics	117
Table 6	Study Sample Characteristics	120
Table 7	Key themes and example quotations from report narratives relating to FRT levels 1 and 2.	125
Table 8	Safety events and concerns in fatigue-related MORs	127
Table 9	Consultation over App development	152
Table 10	Descriptive statistics of participants	166
Table 11	Flight and duty data from pilot’s work schedules	168
Table 12	Biomathematical predictions of main sleep periods prior to flying duties	169
Table 13	Biomathematical predictions of Karolinska Sleepiness Score (KSS) levels at Duty points	170
Table 14	Flying hours associated with high KSS predictions and elevated continual hours of wakefulness	171
Table 15	KSS submissions across the study period	171
Table 16	Comparison of Model 3-6 outputs	176
Table 17	Appendix ii: Practical implications for reporting requirements	267
Table 18	Appendix iii: Main R packages for analysis	268

Chapter 1. Literature Review

1.1 Introduction

The hazard of operator fatigue has long been recognised as a significant potential risk to safe pilot performance and flight safety within commercial aviation operations (Caldwell, 2005, 2012; National Transportation Safety Board, 2019; Wilson et al., 2007). Indeed, over the past forty years over 200 safety recommendations focussed on fatigue have been issued by the National Transportation Safety Bureau (Marcus & Rosekind, 2017) and the management of fatigue risks has remained high on the list of safety-related priorities for many aviation regulatory bodies, including the UK Civil Aviation Authority (Civil Aviation Authority, 2020; Caldwell et al., 2012). As will be discussed in more detail in the following sections, pilot fatigue remains a central safety concern in the aviation industry due to the enduring risk of sleep deprivation that arises from the combination of long working hours, transmeridian travel and rotating shift patterns which interfere with sleep drives (Caldwell et al., 2019; Dawson & McCulloch, 2005). In addition, pilot fatigue has been implicated as either a causal or contributory factor to aircraft crashes and serious incidents on a number of occasions (Drury et al., 2012; National Transportation Safety Board, 1994, 2010; Rosekind et al., 2000; Swiss Aircraft Accident Investigation Bureau, 2001) and as such, high levels of sleepiness have been found to not only erode pilot performance but also endanger the overall safety of flight.

The interdisciplinary literature review for this thesis is structured in the following way. First an overview of the science underpinning sleep and wake processes and the shift work factors influencing on-duty sleepiness is provided. Next laboratory research on the neurobehavioural consequences of sleepiness and applied field studies on the effects of fatigue-related performance decline in aviation are discussed. The safety standards controlling work hours and governing pilot performance are subsequently described, providing relevant industry context for interpreting existing risk exposure levels and safety reporting in commercial aviation. Finally, evidence regarding the burden of fatigue risk exposure across the aviation industry is examined and limitations in both applied field research of in-flight sleepiness rates and the data captured by incident databases are discussed.

Conceptual clarification of ‘Operator fatigue’

The definition of ‘fatigue’ varies not only between the cognitive, sleep and clinical academic literatures, but also in how the construct is described by industry regulations or referred to by transport practitioners (Phillips, 2015). Typically this latter group seek a practical catch-all term that bounds causes and performance effects of ‘fatigue’ within the framework of a safety hazard problem. Such frameworks are often correspondingly somewhat bespoke to the industry context than more precise definitions of distinct but related phenomena established in cognitive and sleep science literatures. However, in recent years, the need to bring these two approaches together has become ever more apparent within the aviation industry (Bendak & Rashid, 2020). The practitioner-held definitions often do not provide sufficient clarity to regulators and safety staff on increasingly complex operational problems. These include how to routinely predict ‘fatigue’ across work patterns, reliably measure fatigue levels of a remote workforce, and develop organisational metrics of operator fatigue risk that meaningfully assess the safety impact of in-flight fatigue rates in the safety-critical context of commercial aviation.

Historically, in common parlance, fatigue has been typically defined as a subjective feeling of tiredness, weariness, lethargy or an unpleasant bodily state, that has been brought about by a summation of different factors, including acute or chronic forms of mental or physical effort which are associated with maintaining current normal levels of behaviour and performance (Hockey, 2012, p12; Phillips, 2015). Like the experience of pain, the symptoms of what most people would term ‘fatigue’ hence can be wide-ranging and indicative of a variety of different causes or conditions, including specific medical syndromes or side effects of medication (Dörr & Nater, 2013; Galland-Decker et al., 2019; Greenberg, 2002). For example, ‘Cancer-related fatigue’ is a phenomenon thought to involve the combination of tiredness, activity reduction, health-related dysfunction and psychological factors (Barsevick et al., 2010) and ‘Chronic Fatigue Syndrome’ is characterised by exhaustion, impaired memory, concentration, headaches, sleep problems as well as musculoskeletal pain (Wyller et al., 2009). This wide variety of definitions presents a challenge for theoretical explanations attempting to explain the brain processes underpinning fatigue or other attempts to characterise the boundaries of this multi-dimensional construct. Indeed, there has been a somewhat longstanding ‘intuitive’ consensus of what fatigue is (Bartley & Chute, 1947; Colman, 2006; Phillips, 2015) but somewhat divergent views as to how it should be measured or operationalized (Aaronson et al., 1999; Gimeno et al., 2006; Musico, 1921; Williamson et al., 2011), or how it can be distinguished empirically from other forms of subjective experience such as anxiety, burnout or boredom (Hockey, 2012; Lal & Craig, 2001) which may share certain features or occur simultaneously. By contrast, it has been easier for researchers to garner consensus on the construct of ‘sleepiness’, which is the brain state associated with instability of wakefulness and the increasing physiological propensity to fall asleep (Durmer & Dinges, 2005). Thus when fatigue is referred to in safety-critical transport settings more as ‘a biological drive for recuperative rest’ (Williamson et al., 2011), this definition relates more clearly to the construct of sleepiness. As will be described in greater detail later in this thesis, ‘sleepiness’ has received much greater academic consensus on its definition, biological causes, measurement and associated performance decrements that are relevant to the transport setting (Åkerstedt et al., 2014; Caldwell, 2012; Cheng & Drake, 2016; Horne & Reyner, 1999). It has therefore emerged at the forefront of research as the key phenomenon of interest with regards to ‘operator fatigue’.

Definition of Pilot Fatigue

The definition and use of the term ‘pilot fatigue’ within the aviation industry reflects a safety hazard concern over the impact that sleep, circadian and cognitive workload factors may have on pilot performance and functioning during flight. As set out by the International Civil Aviation Organisation (ICAO, 2012; 2020), ‘Pilot fatigue’ is functionally defined as:

> *“A physiological state of reduced mental or physical performance capability resulting from sleep loss or extended wakefulness, circadian phase, or workload (mental and/or physical activity) that can impair a crew member’s alertness and ability to safely operate an aircraft or perform safety related duties.”*

This definition is bespoke insofar as its outcome remit refers specifically to alertness decrement but also a context-specific concern regarding pilot in-flight safe performance. The challenge of this definition is that it does not unequivocally state how ‘fatigue’ under the chosen wording should be measured. Arguably, when considering the broader sleep science, cognitive and multi-disciplinary academic literatures, there could be at least four more major drivers of ‘fatigue’ than the ICAO definition includes that could act to compromise safe pilot performance (Figure 1). These include psychological factors such as stress, boredom or motivation, which in a performance setting, may sometimes be referred to as ‘task-driven’ or ‘mental fatigue’ to reflect how an individual copes with the information processing demands of different tasks (Chen et al., 2014; Fard & Lavender, 2019). There are also further modifying factors (medical, physical, genetic, and environmental) which can individually or in combination with other factors influence an individual's alertness and variability in their performance on tasks, that are discussed later in Section 1.2.

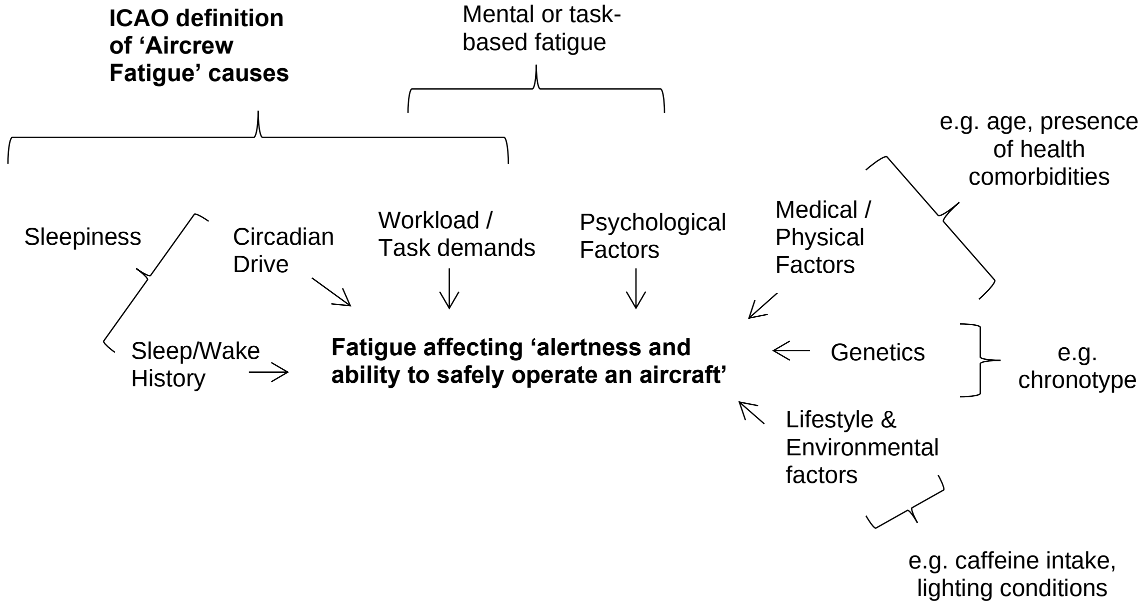


Figure 1. Conceptual clarification of sleepiness and fatigue terms from academic and applied field literatures

In practice, sleep and circadian drives have emerged as the key factors of research interest with regards to 'aircrew fatigue' (ICAO, 2012), and the term is often used interchangeably with 'sleepiness' (the brain state associated with instability of wakefulness) and the increasing physiological propensity to fall asleep (Durmer & Dinges, 2005; Phillips, 2015). For the purposes of this thesis, pilot sleepiness driven by sleep and circadian factors will be the primary focus of the field investigations into the safety hazard of fatigue. However, necessarily for the translation into operational circles, there will be some need within sections to refer to both sleepiness, the scientific construct, and 'pilot/aircrew fatigue' denoting the safety hazard term widely used in aviation studies and safety reports by transport researchers and practitioners.

1.2 Sleep science Background

This section describes a brief overview of the existing literature on sleep science, including the main causes of sleepiness and the shift factors influencing sleepiness and on-duty sleepiness.

Sleep

Whilst its exact purpose is not fully understood, most scientists agree that sleep is a vital physiological function (Carskadon & Dement, 2011; Chaput et al., 2020) that provides ‘rest and recovery from the wear and tear of wakefulness’ (Horne, 1988). Sleep is thought to be important for different types of body restitution such as tissue recovery, immune functioning, thermoregulation, and also for some essential aspects of cognitive processes, such as memory consolidation (Alhola & Polo-Kantola, 2007; Maquet, 2001; Stickgold, 2005). On average, healthy adult humans require between 7 and 8.5 hours of sleep per night to feel well and maintain full cognitive effectiveness (Kripke et al., 2002; Kronholm et al., 2006; Luyster et al., 2012). However, like other physiological parameters, individuals vary in their need for sleep, with some individuals requiring more or less than 8 hours per night (Shneerson, 2000; van Dongen et al., 2012) or showing differences in their trait vulnerability to the effects of sleep loss (van Dongen, 2006; van Dongen et al., 2004). Where individuals do not achieve their required sleep need, they become sleep deprived and their resulting “sleep debt” can accrue in different ways (Alhola & Polo-Kantola, 2007). Sleep debt can be caused by acute total sleep deprivation (TSD) resulting from a single period of continual wakefulness. It can also be caused from limited sleep of typically less than 6 hours per 24 hour period over several days (chronic partial SD), or some combination of both acute and partial forms of sleep deprivation. Outside of laboratory experiments, it is typically chronic partial SD that affects most individuals in their daily lives, particularly for those who do shift work outside the normal day-time hours (Barger et al., 2009; Lockley et al., 2007).

The brain is able to compensate somewhat for sleep loss, as not every hour lost during sleep deprivation is subsequently needed in order for humans to recover normal levels of alertness. Indeed, most sleep loss diminishes stage 2 and rapid eye movement (REM) sleep, whereas slow wave sleep (SWS), the phase thought to be particularly important for recovery, is retained (Åkerstedt, 2003). However, the speed of recovery appears to depend on the manner by which sleep has been lost. For example, following conditions of TSD, the body will typically compensate this loss in the following sleep period by changing the architecture of the sleep, so that one rapidly enters the deep stages of sleep (van Dongen et al., 2003). However, following chronic partial SD, recovery appears to be much slower, potentially taking several days or even a week (Caldwell et al., 2012). The neurobiological reasons for this are not fully understood, but adenosine regulation seems to play a role in differential recovery rates from sleep deprivation (Basheer et al., 2004; Huang et al., 2011; Kim et al., 2015). In acute, total sleep deprivation, the increase in adenosine levels is reversed with subsequent recovery sleep, with no lasting change to the relative receptivity of the basal forebrain to adenosine. By contrast, research with rats has indicated that chronic sleep restriction (of 5 or more consecutive days) is associated with an increase of adenosine receptors, which form in response to increased levels of extracellular adenosine (Kim et al., 2015). When recovery sleep is eventually obtained, such receptors do not disappear automatically, and thus the down regulation to normal levels of extracellular adenosine following chronic sleep deprivation may take several days in a week or possibly longer (Caldwell et al., 2012).

Homeostatic and circadian factors influencing daytime alertness

Alertness levels in humans show a consistent, predictable pattern of peaks and troughs during the 24-hour day-night cycle. This naturally occurring pattern of increased alertness during the day and sleepiness and inactivity at night is the result of an interaction between the diurnal pattern of sleep and wakefulness and the sinusoidal circadian cycle (Åkerstedt, 1995; Åkerstedt et al., 2009). Among the many factors that influence sleepiness, some can be quantified reasonably well in terms of their effects on alertness and neurobehavioural functioning (Åkerstedt & Folkard, 1996; Dawson et al., 2011; McCauley et al., 2013). The duration of prior wakefulness (homeostatic drive) and the time of day (circadian drive) are two such factors.

The homeostatic drive

One’s degree of ‘sleep drive’ or propensity to fall asleep at any given point in time is modulated by endogenous biochemical systems regulating sleep-wake homeostasis. Setting aside circadian influences, in healthy humans the propensity to fall asleep acts as a function of the amount of time elapsed since the last adequate sleep episode (Åkerstedt et al., 2008; Åkerstedt & Folkard, 1996; Borbély, 1982). The resulting effect of sleep homeostasis regulation is fairly intuitive; the longer humans remain awake, the greater the desire and need to sleep, and the greater the likelihood of falling asleep. By contrast, the longer humans are asleep, the more the pressure to sleep decreases, leading to a greater likelihood of waking up. In terms of the underlying biology, sleep-wake processes are controlled and modulated by a set of complex neurochemical pathways and sleep regulating substances. Whilst the full array of neurochemical interactions underpinning the homeostatic drive have yet to be determined, some candidate molecules and neurotransmitters have been implicated (Basheer et al., 2004). What has been established is that endogenous sleep-regulating substances build up in the body's cerebrospinal fluid during waking hours, and rapidly degrade during sleep. The most well established biological marker to show this pattern is adenosine, which is generated as a natural by-product of the breakdown of internal energy stores (Basheer et al., 2004; Porkka-Heiskanen et al., 2002). Adenosine appears to contribute to sleep onset by inhibiting wakefulness-promoting neurons in the basal forebrain. As the number of hours awake increases, brain glycogen reserves and adenosine triphosphate (ATP, molecule that stores energy) levels diminish and cause adenosine levels to accumulate in the basal forebrain (Porkka-Heiskanen et al., 2000). Adenosine levels then rapidly degrade during the following period of sleep, and are gradually replaced by new stores of glycogen (Benington & Craig Heller, 1995). Various studies in animals support a key role of adenosine in the homeostatic sleep drive. In animals, injections of adenosine or similar compounds induce apparently normal sleep (Benington & Craig Heller, 1995; McCarley, 2007) and animals that are sleep deprived show dramatically increased concentrations of adenosine in the brain (Porkka-Heiskanen et al., 2000). Moreover, blocking adenosine's actions in the brain increases alertness (Ribeiro & Sebastio, 2010). In humans the impact of adenosine receptor blocking can be seen in the temporary alertness-boosting effect associated with the consumption of caffeinated products. Caffeine acts as an adenosine antagonist or receptor blocker, which is effective in temporarily diminishing the drive to sleep in humans (Ribeiro & Sebastio, 2010).

Circadian rhythms

The other main processes modulating sleepiness and alertness levels in humans are the circadian rhythms, which fluctuate throughout the day and act as an internal biological clock to regulate the timing of metabolic processes and behavioural functioning (Arendt, 2010). The main internal 'pacemaker' which controls the firing of nerve cells underpinning circadian rhythmicity is located in the suprachiasmatic nuclei (SCN) of the anterior hypothalamus. The SCN is important for determining sleep-wake cycles, brain activity, core body temperature, blood pressure and other biological activities such as hormone production and cell regeneration which appear to coincide with these daily cycles (Hastings et al., 2007). SCN activity produces alerting pulses that interact with homeostatic sleep pressure across the day-night cycle. These alerting pulses peak around 2-3 hours prior to one's habitual bedtime, which counteracts the homeostatic sleep drive that has been continually building throughout the waking hours, and permits wakefulness to continue later into the evening (Dijk & Archer, 2009). Over the course of the evening, the SCN alerting activity decreases, while activity causing the production of melatonin increases, together causing the propensity to sleep to begin to dramatically increase in the early hours of the morning. The SCN or biological clock is entrained to the environment by external cues, or zeitgebers, which aid the natural synchronization of internal rhythms to the Earth's 24-hour rotation cycle (Arendt, 2010; Dijk & Archer, 2009). Such external time cues include the light-dark cycle which plays a prominent role in calibrating the timing of SCN activity. Whilst the circadian rhythm will continue to show predictable patterns of oscillations whether or not zeitgebers are present, the light-dark cycle (in concert with other zeitgebers like the timing of meals etc), acts to entrain the timing of biological rhythms to the local environment (Arendt, 2010; Dijk & Archer, 2009; Roenneberg et al., 2013).

Homeostatic and circadian processes have been subject to extensive research in both biological and bio-mathematical literatures. Experimental sleep deprivation studies suggest that on many aspects of cognition, overall performance declines as a function of time spent awake, and this decline in performance is modulated by circadian rhythm (Durmer & Dinges, 2005; Goel et al., 2013). For the typical day, this means that subjective sleepiness and sleep-driven performance lapses are low across the first 16 hours of wakefulness, but then increase across the habitual night, peaking at around 26 hours awake (Åkerstedt & Wright, 2009). To date, the most useful model for predicting physiological sleepiness and the likely sleep-wake cycle in humans is the two-process model, that mathematically charts the interaction of sleep and circadian factors (Borbély, 1982; Borbély & Achermann, 1992). The homeostatic sleep drive is modelled as a pattern of increasing sleepiness with increasing periods of continual time spent awake, and the recovery of alertness during sleep. The circadian processes are described by the interaction of a pair of sinusoidal waves, the circadian rhythm lasting twenty-four hours, and the ultradian rhythm lasting twelve hours. The aggregation of these processes produces an estimated level of sleepiness at any given moment. The overall effect of this mathematical profile is that humans are alert throughout most of the morning, afternoon and early evening (save for a small dip in the early afternoon), but that this alertness decreases quite rapidly as the night progresses, where the drive from both sleep and circadian factors is towards sleep (Basner et al., 2013). Although there may be some differences in attempts to mathematically chart this profile depending on e.g. presumed light availability (Shen et al., 2006) or chronotype (Kerkhof & van Dongen, 1996), many established models of sleepiness, if not all, appear to be founded on this baseline formulation (Civil Aviation Safety Authority Australia, 2014). While the two-process model is commonly used to make predictions of alertness and sleepiness, its original purpose was to make predictions regarding the timing and duration of sleep (Dawson, 2012). Given a specific pattern of work, the two-process model is capable of making predictions regarding the timing and duration of sleep that an average person would experience, to a reasonable degree of accuracy (Dorrian et al., 2012). Extending beyond this, several biomathematical models that predict sleepiness or fatigue levels during waking hours from sleep estimations have been validated against performance in laboratory, driving, aviation and shift work settings (Åkerstedt et al., 2008; Dawson et al., 2011; Ingre et al., 2014; Kandelaars et al., 2005; van Dongen, 2004). The use of biomathematical models to predict sleep opportunities and approximate on-duty sleepiness has hence become particularly important in safety-critical work environments where operator fatigue risks are elevated by the intrinsic nature of different types of shift patterns.

There are a number of other individual, operational and environmental factors which also may influence sleep and sleepiness levels in any given individual, the most prominent of which are age, genetic differences, medical conditions and lighting. In general, older individuals tend to achieve less sleep which is more disrupted. More specifically, sleep complaints, including sleep fragmentation and other forms of disrupted or truncated sleep normally increase with age (Åkerstedt et al., 2002; Gander et al., 1993), and older individuals tend to have more nocturnal awakenings, and propensity to nap during the day (Rosekind et al., 2000). In the operational context this means that the age of pilots may affect their experience of sleep and recovery from shift work duties (Gander & Signal, 2008). There are not many studies that have investigated the effect of age as contributor to on-duty sleepiness, but in general, the evidence is mixed within the healthy population. Some studies have suggested that older shift workers have been found to encounter greater sleep loss on a given work pattern than younger counterparts (Bonnefond et al., 2006; Gander et al., 1993), and according to some survey work in the aviation industry, younger pilots appear better able to resist the effects of 'fatigue' on performance than older counterparts (Bourgeois-Bougrine et al., 2003). For example, in a survey with 739 airline pilots, younger pilots reported less of an impact of subjective fatigue on flying tasks such as ‘selecting and entering data’, ‘check-lists’, ‘writing official reports’, and ‘flight path monitoring’ than older counterparts (Bourgeois-Bougrine et al., 2003). However, in a more recent questionnaire study of sleepiness prevalence in Portuguese airline pilots (n =1500) where the age range was from 20-60 years, age was not found to be significantly associated with self-report scale measures of fatigue, daytime sleepiness or sleep complaints (Reis et al., 2016). Other studies have found a more counterintuitive significant effect of age associated with lower fatigue or reduced propensity to fall asleep at work (Åkerstedt et al., 2004; Watt, 2000). This might suggest that for certain professions, staff that struggle with fatigue and sleepiness issues might self-select out of fatigue-inducing work with increasing age (Åkerstedt et al., 2004).

Studies using sleep deprivation paradigms have revealed that there is substantial variation between individuals in terms of their response to sleep loss, and to a more limited degree their internal circadian rhythm parameters. For example, it is now well established that, controlling for lifestyle and environmental factors, individuals show reasonably large stable differences in terms of the degree of fatigue and sleepiness experienced, and the cognitive performance vulnerability to both forms of acute and chronic sleep loss (van Dongen, 2006; van Dongen et al., 2011, 2012). When sleep deprived, some individuals are highly vulnerable to neurobehavioral performance decrements, others show greater levels of neurobehavioral resistance to sleep loss, and others show intermediate responses (van Dongen et al., 2004). To date, studies indicate that these phenotypic responses occur as a normal distribution, which suggests the phenotype may be polygenetic (van Dongen et al., 2004). However, more studies are warranted to investigate potential genotypic markers of phenotypic vulnerability to sleep loss and the differential role they might play in response to different types of sleep loss and subsequent performance. Although the internal circadian rhythm is remarkably resistant to rapid change, research has indicated that like sleep need, individuals differ in various endogenous circadian parameters including circadian period, circadian amplitude, and circadian phase (Czeisler et al., 1999; Goel et al., 2013). For example, on average, the human circadian period (or tau) ranges somewhere between about 23.5 -24.5 hours and the genetic basis of this variation is linked with the period or PER gene (Chang et al., 2019; Czeisler et al., 1999; Scheer et al., 2007). However, one of the most well documented of the circadian variations with respect to human functioning and performance is chronotype or circadian phase. Chronotype or morning/evening-type refers to one’s tendency to be an early ‘lark’ or a late ‘owl’ that is endogenously driven (as opposed to lifestyle driven). Morning types fare better (are more alert and function better) in the early hours, but struggle with sleepiness relatively early in the evening, while conversely, evening types struggle in the early hours and do better at the end of the day (Roenneberg, 2012). Intermediate types may fall somewhere inbetween or vary in their preference (Martynhak et al., 2010). In adults, factors such as age and gender also influence morning-eveningness, with some studies suggesting a shift to earlier circadian phase and usual wake time in older individuals (Adan & Almirall, 1991; Duffy & Czeisler, 2002), and on average, there is a suggestion of a greater skew toward earlier circadian phase in women compared with men (Roenneberg et al., 2007). These differences in circadian phase preference (and possibly in circadian period) appear to be enduring traits, with a significant genetic basis. For example, morningness–eveningness is estimated to be about 50% heritable (Barclay et al., 2013). The effect of chronotype on sleepiness and other types of fatigue in work settings has not been extensively studied and the findings are somewhat mixed, in part due to the fact that for some work, staff may be able to engage in mitigating behaviours to minimise their on-duty sleepiness risks, where their intrinsic preferences do not match their work hours. Some studies have shown that in shift work, late chronotypes in particular may sleep less than their early or intermediate chronotype counterparts. For example, evening types have been shown to exhibit higher daytime sleepiness and perform worse in the morning across both cognitive and physical measures compared to morning chronotypes (Facer-Childs et al., 2018), and exhibit shorter main sleep duration prior to morning shift periods, than earlier chronotypes do for night shift periods (Juda et al., 2013; van de Ven et al., 2016). One reason for this might be that shift workers in general find it more difficult to go to sleep early prior to an early shift start, and do not tend to proportionally adjust their bed time in line with early shift start times (Åkerstedt et al., 2010), whereas early types may be able to more easily take advantage of a nap period prior to a night shift. For example, in an observational study of 96 intensive care unit (ICU) nurses that used actigraphy (measurement of wrist movement) to measure sleep patterns, chronotype was not significantly associated with performance (as measured by declines in vigilance or problem solving accuracy) during day or night shifts, although morning chronotypes were more likely to achieve naps and extended sleep opportunities right before starting a night shift (Reinke et al., 2015). However, a recent review of the literature on chronotype and performance in shift work has suggested that in general, there is still a lack of data on the association between chronotypes under variable shift patterns (where work hours may change rapidly over consecutive days or short time periods) and it is not known if tailoring work hours based on chronotype reduces on-duty sleepiness levels in different employment settings (Rosa et al., 2021).

A variety of different medical conditions and medications can interfere with both sleep quality and quantity, consequently contributing to increased daytime sleepiness in individuals (Smolensky et al., 2011).Whilst it is beyond the scope of this thesis to assess the evidence for different medical conditions and medications, some key findings of the major classes of influence and effects are presented in a recent review by Caldwell and colleagues (Caldwell et al., 2019). In general, it is the case that any medication or condition that disrupts sleep is likely to generate subsequent on-duty sleepiness, since sleep disruptions or truncated sleep periods are not adequately restorative (Caldwell et al., 2019). One of the most problematic disorders with respect to sleep loss in shift workers is obstructive sleep apnoea (OSA), which is estimated to have a population prevalence affecting between 9 - 38% of adults (Hartzler, 2014; Senaratna et al., 2017; Williamson et al., 2011). Individuals with OSA suffer from sleep disruptions associated with the inability to breath from the partial blockage of the upper airway, which causes decreased sleep quality and greater daytime sleepiness in affected individuals. Whilst undetected OSA represents a major concern for many shift-working professions, it is worth noting that this disorder is less likely to affect active UK commercial airline pilots, as their strict semi-annual medical certification process would most likely identify this disorder and would assess the pilot as unfit unless the OSA was satisfactorily treated (Civil Aviation Authority, 2021a).

Finally, environmental factors such as ambient lighting, noise and temperature can also act to enhance or degrade sleep quality and maintenance (Cao et al., 2021; Hartzler, 2014; Tembo & Parker, 2009). Dark, quiet and comfortable sleep environments appear to be most conducive to sleep, whereas too much light can disrupt sleep because it is a primary external cue that the brain uses to encourage wakefulness (Tembo & Parker, 2009). An environment that is either too hot or cold will also cause frequent awakenings from sleep due to discomfort, and loud intermittent sounds interrupt sleep as the brain is unable to habituate to these types of noise intrusions (Honkus, 2003). These environmental factors and others may act alone or in combination to affect sleep quantity or quality in pilots, or during wake hours, may act to promote alertness or sleepiness while on duty (Civil Aviation Authority, 2003b; Hartzler, 2014). Whilst the impact of different environmental conditions on on-duty sleepiness has not been extensively studied, most research suggests a particularly important role of lighting conditions as the dominant cue entraining circadian rhythms in applied field settings (Terman & Terman, 2005). Early research in the aviation environment found increased feelings of fatigue may be influenced by dim lighting on the flight deck (Graeber et al., 1990), and more recent exploratory research has suggested that exposure to blue light therapy in a small sample of flight and cabin crew (n =14) was significantly associated with increased alertness levels compared to baseline (pre-intervention levels) (Schoutens et al., 2014). On a larger occupational scale, a recent randomised crossover clinical trial with ICU nurses found that the participants experienced reduced sleepiness during the night shift with high illuminance lighting compared with standard hospital lighting (Griepentrog et al., 2018). Thus whilst at the aggregate level, homeostatic and circadian sleep drives are the predominant factors influencing sleepiness in humans, at the individual level, there may be a number of additional factors that may act to increase or reduce sleep and on-duty sleepiness levels in pilots.

1.3 Shift factors influencing sleepiness

Transport researchers have become increasingly aware of how the scheduling factors intrinsic to continuous operations around the clock may cause sleepiness and operator fatigue-related decrements in performance and increased accident risk (Williamson et al., 2011). As set out in the previous section, two primary drivers of alertness at any given point in time are recent sleep and the body clock. If recent sleep is insufficient or the body clock is at a ‘low point’ (as is the case with night work) or desynchronized (as is the case with jet lag or rapid changes in shift start times), sleepiness will be exacerbated (Åkerstedt & Folkard, 1996; Åkerstedt & Wright, 2009; Caldwell, 2012). Thus the timing and duration of work shifts and rest time between duties are amongst the most prominent influencers of operators’ sleep and on-duty sleepiness (Sallinen & Hublin, 2015; Sallinen & Kecklund, 2010). Various authors (Åkerstedt & Wright, 2009; National Research Council, 2011; Phillips et al., 2017) in summarising these findings have suggested the common overarching shift factors influencing operator on-duty sleepiness are:

- The time of day of the transport operation
- The period of continual wakefulness
- The length and quality of prior sleep
- The length of working duty hours

Research on shift work patterns that relate to these factors has predominantly had a focus on night shift work as presenting the greatest risk to operator sleepiness and safe performance.

Night Shift Work

It is now well documented that working during the night (circadian nadir period) and trying to sleep during the day is a cause of sleep loss in night shift workers (Åkerstedt, 2003; Pires et al., 2009). In comparison to normal healthy adults, who show roughly between 7-8 hours sleep each night (Groeger et al., 2004; Rajaratnam et al., 2004) night shift workers show much shorter and poorer quality sleep during the day (Åkerstedt, 2003; Arendt et al., 2006; Axelsson et al., 2004; Kecklund & Åkerstedt, 1995). Indeed, it has long been established that differences in sleep length appear to be related to the timing of shift worked, with workers on afternoon/evening shifts showing the longest sleep, those working during typical day shifts slightly less, and night shift workers showing the least amount of sleep (Åkerstedt & Torsvall, 1981; Tepas et al., 1985; Williamson & Sanderson, 1986). More recent studies have also shown that shifts requiring wake-ups during the biological night time also truncate sleep substantially (Ingre et al., 2004, 2008). Objective assessment via electroencephalography (EEG) and polysomnographic monitoring of rotating shift workers in train, hospital, factory, and long-haul truck driving industries furthermore indicates that sleep taken during the day is 1–4 hours shorter than night sleep (Lockley et al., 2004; Mitler et al., 1997; Tilley et al., 1982; Torsvall et al., 1989). Night shifts are particularly problematic for sleep loss as people do not tend to adjust to day sleep even across a series of night shifts (Åkerstedt & Wright, 2009; Dahlgren, 1981; Wright et al., 2013).

A wealth of shift work studies have indicated high levels of on-duty sleepiness in partial and permanent night workers during their duties, as measured by both subjective sleep reports and physiological measures (e.g. Åkerstedt & Wright, 2009; Bjorvatn et al., 2006; Haidarimoghadam et al., 2017; Paley & Tepas, 1994; Wilson et al., 2019). It should be noted that subjective measures used in such applied field studies (described further in Chapter 2.3) range between established scale metrics (such as the Stanford Sleepiness Scale, SSS; Epsworth scale, ESS; and Karolinska sleepiness scale, KSS), reports of falling asleep or sleep-related behaviour and more vague or one-off measures of sleepiness and operator fatigue. Thus summarising the subjective data across all shift work studies lacks some precision, and direct comparisons between rates are often not possible. Nevertheless, reviews of the literature suggest that the overwhelming majority of shift workers report sleepiness in connection with night-shift work, and rarely report elevated sleepiness during normal consistent day shifts (Åkerstedt & Wright, 2009; Arendt, 2010). In many studies a majority of shift workers admit to having involuntarily fallen asleep on the night shift, with between 10-20% reporting that they regularly fall asleep regularly during nocturnal work (e.g. Åkerstedt, Ingre, et al., 2008; Luna et al., 1997). Studies that have sought to obtain a detailed picture of subjective sleepiness in shift work do so by repeatedly sampling participants across their shifts and during their time off (Åkerstedt & Wright, 2009). Research efforts employing this technique with a mixture of established scales (SSS, KSS) and simple reaction time tasks have indicated that moderate to high levels of sleepiness are evidenced during the night shifts, compared with no evidence of sleepiness at all during day shifts (Härmä et al., 2002; Lowden et al., 1998).

Objective measures of sleepiness also reveal high levels of on-duty sleepiness and sleep in night workers, and point to particularly severe effects (such as involuntary sleep) as being likely to occur in the early morning hours. Such measures are often easier to compare across different study designs. As will be set out in greater detail in the Chapter 2, in general, EEG is recognised as one of the gold standard measures of physiological sleepiness. Alpha activity (8 - 12 Hz) is an EEG pattern associated with relaxed wakefulness and a relatively early sign of increased sleepiness, and theta (4 - 7 Hz) is an EEG pattern associated with more severe sleepiness (Liu et al., 2009). Slow eye movements (SEMs) usually accompany the transition from wakefulness to sleep (de Gennaro et al., 2000; Ogilvie et al., 1988). In an EEG study of night shift train drivers, it was found that one quarter of the drivers displayed marked increases in sleepiness-related alpha and theta activity, alongside slow eye movements towards the early morning part of their shifts (Torsvall & Åkerstedt, 1987). This activity correlated with subjective ratings of sleepiness (KSS) (r =.74.) Moreover, analysis of the performance and physiological data together revealed that instances of prominent degradations in driver performance, such as missing or driving through a red light, occurred during these bursts of SEMs and increased alpha and theta activity. Similar findings have been shown in long haul truck drivers (Kecklund & Åkerstedt, 1993; Mitler et al., 1997). Other EEG research on night shift workers has shown that involuntary sleep and attentional failures due to sleep-related activity tend to occur in the second half of the night shift (Landrigan et al., 2004; Torsvall et al., 1989), and that whilst subjects are generally aware of rising subjective levels of sleepiness, they may often be unaware of having fallen asleep (Reyner & Horne, 1998; Torsvall et al., 1989). Research efforts using operational simulations have also shown increased sleep-related EEG activity during the night time, particularly in the driving context (Åkerstedt et al., 2005; Gillberg et al., 2003; Reyner & Horne, 1998). In these studies, increases in alpha and theta activity were accompanied by large increases in subjective sleepiness.

Similar findings indicating the increased severity and frequency of sleepiness in night-time operations have been reported in the aviation sector. Research has shown that on-duty microsleeps (short episodes of sleep lasting only a few seconds) and longer forms of involuntary sleep are particularly problematic during night time hours compared with flights during the daytime hours (Rosekind et al., 1994). Wright and McGown (2001) reported that whilst pilot sleepiness increases as a function of increased flight duration in both night and day duties, occurrences of EEG-determined sleep episodes were more frequent on flights with late night departures than those departing earlier in the day. Sleepiness appears to accrue substantially over two consecutive night flights, with added homeostatic pressure arising from daytime sleep reductions during layover periods (prior to the return trip) of up to 2 hours (Åkerstedt & Gillberg, 1982; Gundel et al., 1995). A more recent study found that for both LH and SH operations, flight duty periods that overlapped the early (00.00-03.00) and late (03.01-06.00) parts of the night were most consistently associated with reduced sleep efficiency and increased subjective sleepiness (KSS) levels (Sallinen et al., 2017). Flight simulation studies have also suggested that pilots experience sleep episodes throughout uneventful night time simulated flying (Neri et al., 2002), and sleep events and lapses are up to nine times more likely at night than during the day time (Caldwell, 2012; Dinges, 1990; Neri et al., 2002). With regards to military pilot performance, research has also shown that flight-control deviations occur more frequently and with greater severity when the flight overlaps the biological night-times of the flight crew (Caldwell, 2005). Indeed, in one night time simulator study, the greatest number of performance errors were found to be made during the early morning nadir period, particularly at points where involuntary sleep lapses were most prominent (Caldwell, 2005; Moore-Ede, 1993). Commercial pilots themselves also report greater subjective fatigue levels during night flights, and in one questionnaire study using exploratory analyses with 1500 Portuguese pilots, night shifts were found to have a statistically significant added risk for higher levels of subjective fatigue (OR = 1.272) (Reis et al., 2016).

Shifts that encroach the biological night appear to be problematic in terms of sleep loss and subsequent sleepiness on the job because of the inability of the circadian system to quickly or fully adapt to a different sleep-wake schedule, and because sleep loss will tend to extend the hours of continual wakefulness in the shift worker (Åkerstedt, 2003). The circadian rhythm shows substantial inertia, and thus adapts slowly, if at all to rapid transitions between shift schedules or waking hours (Kuhn, 2001). Even following rapid changes in time cues, the daily phase shift in circadian timing rarely exceeds more than 1-1.5 hours on average without interventions (Arendt, 2010). For rotational shifts that switch between early, late and night hours, this may mean that there is not enough time for the circadian rhythm to adapt in the short-term. However, as mentioned previously, even longer term night shift workers do not appear to adapt their circadian rhythm to their nocturnal working schedule (Folkard, 2008). Full circadian adaptation may require the maintenance of the same night activity and day sleep pattern on days off, as research has suggested that partial shifts or total reversals in circadian timing are related to the degree to which workers are exposed to natural light (Arendt, 2010). Social cues from normal daily activity with friends and family also appear to have a role in preventing full adaptation of one’s circadian rhythm to night shifts, albeit much weaker than light (Mistlberger & Skene, 2004; Patton & Mistlberger, 2013). This is because social cues typically follow a diurnal pattern, and it can be difficult outside of laboratory or unusual environments to reduce the influence of such cues or ‘zeitgebers’ (Arendt, 2010). Where such a reduction of zeitgeber influences is possible, often in very exceptional environments and circumstances (such as base workers in the Antarctic), partial or complete circadian adaptation to permanent night work has been shown to be possible, but not guaranteed (Arendt & Middleton, 2018; Boudreau et al., 2013).

Shift work requiring early wake up times

Research has revealed that early morning shifts requiring wake-up times prior to 07.00 are also problematic for sleep loss prior to the shift, daytime sleepiness and feelings of being refreshed by sleep (Åkerstedt & Wright, 2009; Ingre et al., 2004, 2008). Indeed, there is some evidence to suggest that sleep achieved by workers before early morning shifts can be shorter and more disrupted even than before night shifts (Folkard & Barton, 1993). EEG studies suggest this shortened sleep (2-4 hours less than basal sleep need) affects predominantly stage 2 and REM sleep (Åkerstedt, 1995; Elmenhorst et al., 2008). In particular, it seems that rising between 04:00 and 06:00 is strongly associated with increased sleepiness during the rest of the day, and prior sleep is reduced where duties start prior to 09:00h (Ingre et al., 2008; Kecklund & Åkerstedt, 1993; Onninen et al., 2020). One experimental field study in the train driving sector found a statistical relationship between morning shift-start time and the associated sleep length the prior night in irregular shift work. Ingre et al. (2008) found that for shifts starting at 03.00-12.00, there was a corresponding curvilinear relation between shift-start time and sleep duration. More specifically, there was a near linear increase in sleep length for those shifts starting between 04.50 and 09.00 of approximately 0.7 h for every 1 hour the shift was delayed. Shifts beginning between 04:30 – 07.00h were seen to have a large detrimental effect in reducing the sleep length achieved by train operators to below that of 7 hours. The shortest sleeps of around 5 hours were found for shifts that required waking up before 04:30. Other field studies with shift-workers have shown similar findings with respect to early duty start times and sleep loss (Bostock & Steptoe, 2013; Flynn-Evans et al., 2018; Spencer & Robertson, 2000) which together suggest increasing sleep loss for every hour that duty periods start prior to 09.00.

Roach et al. (2012) conducted a similar study in the aviation sector using a combination of objective data from sleep diaries and activity monitors, and subjective data (ratings on the Samn-Perelli fatigue scale) to examine the sleep and fatigue levels in short-haul (SH) pilots. Corroborating the findings above, pilots showed shorter sleep and higher levels of fatigue when duty periods began earlier in the morning compared with later duty start times. More specifically, for duty start times between 09.00-10.00, pilots obtained an average of 6.7 hours of sleep in the prior 12 hours to their duty, compared with duty periods that began between 04.00-5.00, where pilots obtained on average just 5.5 hours of sleep in the prior 12 hours. This study showed a similar but milder relationship between duty start time and sleep length, indicating that approximately 15 - 30 minutes of sleep is lost for every hour that duty periods are advanced prior to 0900 (Roach et al., 2012). Survey findings with short-haul pilots furthermore suggest that duty periods with early-morning starts are considered to be one of the key contributors to elevated levels of on-duty subjective fatigue (Bourgeois-Bougrine et al., 2003; Jackson & Earl, 2006), and where pilot duties consist of several consecutive early duty days, pilot fatigue levels appear to accumulate across the block of duties (Spencer & Robertson, 2002). As an operational consequence, the performance of pilots who are rostered long duty durations coupled with early duty start times may be more at risk of impairment during the approach and landing phases of flight (Bendak & Rashid, 2020).

Early morning shifts appear to be problematic in terms of sleep loss for a couple of reasons.

First, early shift starts are likely to be difficult for individuals because they will typically require forced early wake up times that coincide with the circadian nadir period, which is ‘protective’ against sleep termination in healthy humans (Åkerstedt, 2003). This difficulty in awakening during the nadir phase is further likely to be compounded by that fact that operators may not have had much sleep by the time they need to wake up. Alerting circadian influences in the evening make it difficult for individuals to phase advance their bedtimes fully, thus reducing the amount of sleep typically achieved prior to the early awakening (Czeisler et al., 1980). Indeed, pilots report that they are not able to get to sleep earlier in the evening to compensate for the early morning duty start times (Samel et al., 2004; Spencer & Montgomery, 1997). There may also be a number of social or environmental cues that cause sleep to be reduced prior to early morning shifts. For example, some authors have suggested that the actual anticipation of the difficulty of awakening may also be a key factor for early shift workers, particularly in rotating shift sectors, affecting not only sleep length but also the amount of slow wave sleep obtained by workers (Kecklund et al., 1997).

Duration of a work day

Research has suggested long duty hours can impact on-duty sleepiness and performance by restricting the amount of off-duty time available for non-work activities (relaxing, exercising, eating, socialising) and the length of sleep operators achieve. Research from the US Department of Transportation (2000) for example, suggests that workdays as long as 10 to 14 hours provide insufficient time for any other activities beyond eating and sleeping, and that this kind of lifestyle may in the long term have an adverse effect on the worker’s accumulating sleep loss, as well as physical and psychological health (Amundsen et al., 2003). In the aviation sector extended flight duty periods are typically associated with long-haul international flights, but long duty days requiring extended periods of wakefulness are also possible in the short haul sector (Bourgeois-Bougrine et al., 2003; FAA: Flight Crew Member Duty and Rest Requirements, 2011; Gawron, 2016; Hartzler, 2014). Studies have shown that pilots do report increasing fatigue levels across the duration of the flight (Gundel et al., 1995; Petrilli et al., 2006) perhaps indicating a combination of time-on-task effect (prolonged period of information processing related to flight tasks) and homeostatic drives. Rosters that contain flight duty periods exceeding 8 hours are thought to present greater risks to pilot fatigue as pilot duty hours may go substantially beyond their flying hours by at least 2-3 hours (Gander et al., 1998; Rosekind et al., 1995). For instance, a 9 hour flight may involve a 12 hour or more duty period, to which additional discretionary periods of time may be added if the operation requires this[1]. Such extended duty times hence have bearing on not only the period of continued wakefulness and on-duty sleepiness experienced by pilots but also can reduce the amount of time available for sleep periods (Bendak & Rashid, 2020), and may have deleterious effects in the long run that go beyond chronic sleep loss. Indeed, one study found that pilots who reported frequently flying into their ‘discretion’ hour showed poorer self- ratings in terms of physical and psychological health and higher subjective fatigue (Jackson & Earl, 2006).

Irregular or disruptive work hours

Irregular shift patterns, and the speed and direction of the rotation between different shift duties have also been linked with sleep loss and performance deterioration (Flynn-Evans et al., 2018). In general, individuals who work irregular hours usually achieve less sleep prior to duty, with some studies providing estimates of 1-2 hours less than the typical basal requirement for most people of 8 hours per day (Folkard et al., 2005; Lac & Chamoux, 2004). Research has indicated that rapid shift rotations within a single week are associated with greater sleep loss in workers compared with slower rotations (across several weeks) (Pilcher et al., 2000) with rapid counter clockwise rotations appearing the most problematic in terms of sleep loss (Akersted & Wright, 2009; Pilcher et al., 2000). By contrast clockwise rotations are less disruptive to workers as the internal circadian drive shows a natural tendency to permit phase shifting towards later hours (Åkerstedt & Wright, 2009). The biological explanation for increased sleepiness under varying duty start times is that there is a mismatch between the internal body clock (which is resistant to rapid change) and disruption of the sleep/wake cycle. This circadian desynchrony means that there is circadian and homeostatic pressure to sleep at times that might not match the duty or rest times or the new environmental cues (as is the case in jet lag).

Schedule-related causes of sleepiness in long- and short-haul flights

A number of studies conducted within the aviation industry have produced similar findings with respect to the common sleep-wake and circadian determinants identified in shift work studies (e.g. Gander et al., 2015; Powell et al., 2008; Powell et al., 2007; Roach et al., 2012;

---

[1] For a summary description of the work elements of a pilot’s duty day, see Appendix i.

Samel et al., 2004). Like other forms of shift work, the timing and duration of pilot duty hours intrinsically come into conflict with homeostatic and circadian drive aspects of human functioning. Hence many features of shift work - early start times, extended work periods, truncated recovery time periods between duties, night work through the window of circadian low, daytime sleep periods and day-to-night or night-to-day transitions across consecutive work periods – can act alone or in combination to increase pre-duty sleep loss and on-duty sleepiness in pilots (Caldwell, 2012; Caldwell et al., 2009; Wingelaar-Jagt et al., 2021). Moreover, the predominant factors contributing to 'pilot fatigue' appear to primarily related to sleep loss for both long-haul (LH, >6 hours) and short-haul (SH <6 hours) flights (Bourgeois-Bougrine et al., 2003; Gander et al., 2015) and duty timing (Reis et al., 2016). Concurring with laboratory sleep deprivation studies (Van Dongen et al., 2003), more sleep in the 24 hours prior to the start of duty in pilots has been found to independently predict lower sleepiness and fatigue ratings and quicker vigilant attention responses (Gander et al., 2015). Moreover, the most recent study of aircrew fatigue across Europe found that increased on-duty sleepiness levels in pilots were predicted by prior sleep and the proportion of the flight duty covering the night time period (otherwise referred to as window of circadian low (WOCL) encroachment between 02:00-05.59 hours) (Sallinen et al. 2020). However, it should be noted that this study solely focussed the assessment of long night duties and schedules that encroach the WOCL and did not assess other roster-related variables. In addition to these common shift work factors, commercial flight crews may also have rosters that cause circadian rhythm desynchrony from the crossing of multiple time zones (jet lag) and rapid rotations of duty times, or face a number of other environmental factors or work pressures from the cognitive demands of the piloting role that may affect their individual fatigue levels (Caldwell, 2012; Gander et al., 2013; Hartzler, 2014). Commercial rosters also include a combination of other unique job specific factors such as the type of flight, the duration of layover and the number of flight segments within a flight duty period (Dorrian et al., 2012; Honn et al., 2016; Powell et al., 2007; Sallinen et al., 2017). With respect to these schedule factors, there are some differences between LH and SH operations.

Long Haul Operations

Within long haul (LH) flights, pilots report that the combination of night flights, jet lag (circadian rhythm disruption) and insufficient recovery are the most important factors affecting pilot fatigue levels (Bourgeois-Bougrine et al., 2003; Graeber et al., 1990; Harris et al., 2001; Samel & Wegmann, 1988; Venus, 2021). Indeed, the scheduling requirements of many LH operations require pilots to operate at times that overlap their biological night time, and to achieve down route or 'layover' rest during their biological day within a relatively short period of time. Bougeois-Bougrine and colleagues provide the example of two successive nights to and from Paris and New York within 48 hours, that include a short layover of 22 hours. In this trip pairing, following arrival pilots obtain sleep at a point that corresponds to their biological day, restricting both the quality and quantity achieved. This sleep is followed typically by an extended period of wakefulness prior to the following departure, which leads to increased subjective fatigue experienced by the flight crew during the returning night flight back (Bougeois-Bougrine et al., 2003).

Various studies have indicated that quantity and quality of sleep achieved by LH flight crew is reduced during layovers (Eriksen & Åkerstedt, 2006; Lamond et al., 2006; Petrilli et al., 2006), suggesting a particular problem for on-duty sleepiness in LH flights may lie in the return duties following a layover. A recent study for example, found that despite having more time available for sleep, long haul pilots slept between 5 to 6 h per 24-h period during their layover as assessed by sleep tracking devices and sleep diaries (Devine et al., 2021). Whilst there may be a number of reasons for this, some authors have suggested that the quality and quantity of sleep achieved down route largely depends on the combination of how closely the individual's likely sleep propensity is able to coincide with local night (Gander et al., 2014), the length of time spend down route, and the direction of travel (eastward or westward) (CAA, 2005). It should be mentioned that another factor moderating the amount of sleep pressure experienced in LH crew during duty is the fact that some schedules where flight hours exceed 9 hours are usually flown by 'augmented' or 'heavy' crews of 3 members (CAA, 2005; Gander et al., 2011; Signal et al., 2013). With a 3-person crew, usually each pilot will be apportioned a third of the available cruise period to use for in-flight rest in a bunk or a non-flying seat. From an operational standpoint, during heavy operations, the quantity of in-flight sleep achieved is particularly important as pilots have to perform the descent and landing (the most demanding phase of flight) at the end of a long work period (Roach et al., 2010). Presence of an in-flight nap opportunity however, does not assure that pilots can always achieve sleep during that period. Objective assessments suggest that if achieved, the quality tends to be poorer with more frequent awakenings than on ground (Signal et al., 2013), and rest opportunities apportioned earlier in a flight result in less sleep than later rest opportunities (Pascoe et al., 1995).

Short haul operations

Given the notable influence of extended night flights on operator on-duty sleepiness, there has been considerable focus by most field-based studies on LH transmeridian operations (Roach et al., 2012). However, increasingly researchers have recognised that the combination of scheduling factors of irregular working hours, long duty days and multiple sectors in SH operations may leave pilots at risk from on-duty sleepiness and fatigue, and that the impact to safe flight may be greater than expected. SH flights even for long duty days (>11 hours) are operated by 2 flight crew, and often there are few if any opportunities to leave the cockpit to rest at any stage during flight. Furthermore, most SH operations involve multiple flight sectors, involving more take-offs and landings within each duty period. The operational importance of this is that these phases of flight are considered to be the most safety-critical (Roach et al., 2012). However, they are also considered to be the most intensive with respect to workload. Surveys have suggested 'aircrew fatigue' is a frequently encountered problem in short-haul commercial pilots (Bourgeois-Bougrine et al., 2003; Jackson & Earl 2006), and that major causes of on-duty fatigue perceived by SH pilots are long duty periods, successive early start times and multiple flights within the duty period (otherwise termed as 'sectors') (Bourgeois-Bougrine et al., 2003; CAA, 2005; Gawron, 2016; Jackson & Earl, 2006; Powell et al., 2007). Indeed, Bourgeois-Bougrine and colleagues (2003) found that subjective fatigue level (as assessed by analogue scale) was affected by the combined effect of flight schedule time and the number of sectors within the duty period. However it is worth noting that in most operations, the number of sectors is closely related to total duty length. These authors furthermore found that subjective fatigue was higher in successive morning duties than in alternating morning and afternoon duties, suggesting a cumulative impact of sleep loss associated with consecutive early duties. Powell and colleagues (Powell et al., 2007) focussed on determining the factors predicting SH pilot fatigue ratings at the end of their last duty sector. In this study, length of duty, time of day, the number of sectors flown and airport of departure all significantly influenced pilot fatigue levels. Within this, duty length and number of sectors (which the authors note were closely-related) increased fatigue levels in a linear manner. By contrast time of day was seen to have a weaker influence for the SH pilots. Further suggestive evidence of an effect of flight sectors on pilot fatigue has come from a flight simulator study that used a randomised cross-over study design with a small sample of regional airline pilots (n= 24) to investigate the manipulation of sectors within a duty day. Honn and colleagues found that both objective (PVT performance) and subjective ratings of fatigue (Samn-Perelli ratings) built across the nine hour workday and were significantly greater in the condition with multiple sectors (n= 5) compared with single sector (Honn et al., 2016).

In summary, there are a number of common schedule factors that affect sleep and subsequent alertness in pilots. These include night work through the window of circadian low, early start times, extended work periods, insufficient time off between work periods, insufficient recovery time between consecutive work periods, number of consecutive work periods, daytime sleep periods, and day-to-night or night-to-day transitions. In addition to these factors, commercial flight crews may also have rosters that cause circadian rhythm desynchrony from the crossing of multiple time zones, or face a number of other environmental factors that may affect fatigue levels (Caldwell, 2005; Caldwell et al., 2019). The majority of aviation studies assessing roster-related variables to date have focused on night flying, specific routes, trip pairings or particular phases of flight over short time frames and a limited number of airlines. Hence there is a considerable gap in the literature relating to not only the on-duty sleepiness associated with broader array of roster patterns, but also over longer time periods. This means the likely sleep-wake patterns associated with typical pilot duty schedules in recent years are not known, and few peer-reviewed published studies have sought to investigate the impact of roster variables across a large number of different airlines. Against this context it has been suggested that real world investigations into elevated fatigue risks in safety-critical operations should begin by assessing scheduling practices for insufficient sleep opportunities afforded by work and extended time on duty (Dawson & McCulloch, 2005).

1.4 Consequences of sleepiness

Sleep propensity, microsleeps and wake-state instability

From a biological perspective, it is well documented that one's propensity to fall asleep increases as a function of prior wakefulness, as evidenced via EEG recordings of changes in the power spectrum of non-rapid eye movement sleep (NREMS) (Achermann & Borbély, 2011; Borbély et al., 1981). In particular, NREMS of frequencies between 0.25 to 7 Hz appear to increase monotonically with extended levels of wakefulness (Achermann & Borbély, 2011; Cajochen et al., 1995), although these signatures may differ considerably between individuals (Tarokh et al., 2015). Sleep deprivation reduces the latencies between lighter stages of NREM sleep to deeper slow wave sleep (Dinges, 1986). In practice this means that following a night without sleep, the time taken during the day for a healthy adult to fall asleep decreases to less than a minute or two on average, and the time taken from sleep onset to slow wave sleep is halved (Dinges, 1986; Durmer & Dinges, 2005). Extensive experimental research conducted using standardised measures of sleep latencies such as the multiple sleep latency tests (MSLT) which assesses tendency to fall asleep, and maintenance of wakefulness test (MWT) which indexes the resistance to sleep, have shown that sleep latencies reduce in response to sleep deprivation (Goel et al., 2009). Thus whether the context for participants is to fall asleep or stave off sleep, sleep deprivation appears to significantly reduce the time to transition from waking to sleeping. Increasing levels of sleep deprivation are accompanied by transient episodes of sleep-like activity, termed ‘microsleeps’ which begin to intrude into wakefulness. To begin with, microsleeps may be quite brief, only detectable in certain sensitive tests such as the Psychomotor Vigilance Test, or via EEG (Dorrian et al., 2005; Lim & Dinges, 2008). Although brief, such microsleeps can be functionally important, causing attentional lapses during tasks that demand vigilant attention (Torsvall & Åkerstedt, 1987). However, with increasing time awake, lapses increase in frequency and duration, ultimately progressing to the point of uncontrolled sleep attacks, which signal the full transition from the waking state (Doran et al., 2001). Both acute and chronic forms of sleep deprivation can produce a high rate of lapsing, that ultimately progresses to full and sustained sleep onset even during goal-directed behavior, such as driving a car (Dinges, 1995; Lim & Dinges, 2008). Research using sophisticated functional magnetic resonance imaging (fMRI) techniques has provided further insights on the impact of these lapses on visual processing. Chee and colleagues demonstrated that when sleep deprived, participants reveal a significant decline in visual task-related activation in the extrastriate cortex (primarily involved in visual sensory processing), and that this reduced activation is particularly evident during lapses (Chee et al., 2008). The same research group were able to show a correlation between this reduced activation within visual processing regions and poorer performance on visual memory and perceptual load tasks (Chuah & Chee, 2008). Hence, such neuroimaging studies have begun to suggest that even basic visual attention and perception processes are compromised during sleep-driven lapses.

Taken together, these findings suggest that one of the most rudimentary and indeed hallmark ways in which sleepiness can impact operator performance is via functional lapses, which may result in errors of omission (failure to respond in a timely manner to a stimulus) or longer episodes involuntary sleep. However, lapsing is not the only consequence of rising sleepiness levels. Increasing response times or errors of commission in cognitive tasks can occur in non-lapse periods (Dorrian et al., 2005; Kjellberg, 1977) suggesting that as sleep debt accrues, performance on cognitive tasks becomes increasingly variable even independently of lapses. Some cognitive researchers have conceptualised this sleep-driven effect on alertness and cognition as wake ‘state instability’(Doran et al., 2001; Lim & Dinges, 2008). According to this theory, alertness and performance variability depends on two competing neurobiological systems mediating sleep initiation and wake maintenance. There is a top-down conscious drive to maintain alertness (where subjects are motivated), which originates in the rostral brain areas, and a bottom up drive from both central and caudal areas to initiate sleep processes. It is posited that the interaction of these two drives that causes heightened ‘moment to moment variability’ in cognitive functioning and performance (Doran et al., 2001). Thus it is argued that wake-state instability is evident in both the neurobehavioural signs of slow eyelid closures, microsleeps, brief lapses and errors of omission in sleep deprived subjects (Goel et al., 2009; Mallis et al., 1998), and a compensatory effort to maintain performance may be reflected in sleep deprived subjects by the increased performance variability, such as errors of commission on cognitive tasks (Lim & Dinges, 2008). As will be explored further in the next section, experimental research findings suggest that top-down control processes in motivated sleep-deprived subjects can exert a substantial compensatory effect in preserving performance and masking the effects of sleep loss (Horne & Pettitt, 1985; Tilley & Brown, 1992). Notwithstanding this, due to the considerable biological need for humans to sleep, even extreme efforts to maintain wakefulness cannot ultimately stave off intrusions of sleep and sleep initiation (Goel et al., 2009). This has been evidenced not only in the laboratory setting, where under TSD conditions subjects report ‘semi-dreaming’ or disengaging during cognitive tasks (Dinges, 1990a; Patrick & Gilbert, 1896) but also in individuals who ultimately fall asleep in real world settings such as when driving vehicles (Horne & Reyner, 1995; Horne & Reyner, 1999).

Cognitive performance decline

A wealth of empirical evidence has revealed that sleep deprivation (SD) adversely affects human performance on a large range of neurocognitive and motor tasks (Alhola & Polo-Kantola, 2007; Basner et al., 2013; Durmer & Dinges, 2005; Goel, 2017; Goel et al., 2009;

Harrison & Horne, 2000). Experimental and modelling studies of the effects of sleep deprivation (SD) on cognitive performance have revealed that on many aspects of cognition, overall performance declines as a function of time spent awake, and this decline in performance is modulated by circadian rhythm (Durmer & Dinges, 2005; van Dongen & Dinges, 2005). Performance on various cognitive tasks decreases as sleep debt accumulates with extended periods of wakefulness, in a manner that appears 'dose dependent' (van Dongen, Maislin, et al., 2003). This means that, setting aside any time of day circadian influences, performance typically worsens with increasing levels of sleep deprivation. Importantly, these effects are not limited to total forms of SD; moderate sleep restriction (of for example, less than 7 hours per night over 4 or more days) can produce similar cognitive decrements to those evidenced following severe acute total sleep deprivation (Dinges et al., 1997; Philip et al., 2012; van Dongen, Maislin, et al., 2003). Repeated days of sleep deprivation between 3-6 hours time in bed increases daytime sleep propensity (Belenky et al., 2003), speed and accuracy declines in performance on working memory tasks (van Dongen et al., 2003), and increased lapse rate on the psychomotor vigilance test (Dinges et al., 1997; van Dongen et al., 2003). To date, one of the best controlled dose-response experiments on chronic sleep loss was conducted by van Dongen and colleagues, who assessed the effects of TSD following 1, 2, and 3 nights without sleep, and compared these with the effects of restricted sleep of 4, 6, or 8 hours across a two week period (van Dongen et al., 2003). On each day of the assessment period, performance on the psychomotor vigilance test, a working memory task and 'cognitive through put' task was assessed every 2 hours, alongside subjective sleepiness and EEG recordings of objective sleepiness levels. As might be intuitively expected, three days of TSD was found to produce significantly larger decrements in performance than any of the three chronic partial SD conditions. However, following 2 weeks of sleep restriction to 4 hours per night, performance on all of the cognitive tasks was found to be equivalent to that seen following 2 nights of TSD. Likewise, restriction of 6 hours per night across 2 weeks produced cognitive deficits equivalent to those evidenced following a single night of TSD. For both of these sleep restriction conditions, cognitive decrements were seen to accumulate each day across the study period in a near linear manner. Interestingly, the subjective ratings of sleepiness did not quite parallel the cognitive performance declines of subjects, instead showing much more subtle increases across the study period. The authors suggest these findings may mean that as subjects become progressively more sleep deprived, there may be an escalating dissociation between how sleepy they feel, and how well they are able to perform on cognitive tasks. Whilst cognitive performance decrements accumulated much more rapidly in the TSD conditions, this study did not find a linear relationship between total sleep debt (total amount of sleep lost) and the effect size of the cognitive performance decrements. The authors instead have suggested their findings may point more to a critical length of time awake within each circadian cycle, estimated statistically at 15.68 hours, beyond which neurobehavioural decrements begin to occur.

Beyond these broader findings, the detrimental impact of total and partial forms of SD on neurocognitive performance appears far from straightforward. A range of studies have revealed little to no effect of acute SD on performance within some cognitive domains (Binks et al., 1999; Lim & Dinges, 2010; Quigley et al., 2000), and even well controlled studies such as that of Van Dongen et al. (2003) have not provided detailed insights on the relative magnitude of negative effects on cognition that SD exerts (Lim & Dinges, 2010). These gaps in knowledge have prompted evolving theoretical interpretations of how these data might be explained, and debates on which cognitive theories together provide greatest predictive power. In addition, some studies have revealed important differences in the form and characterisation that such performance decrements can take, particularly across different experimental tasks, SD manipulations, cognitive domains and individuals (Van Dongen et al., 2004). However, such intricacies relating to more detailed performance outcomes, including response time and accuracy distributions or variations in performance occurring during simultaneous cognitive components, (such as decision processes), have only recently begun to be teased apart, as most SD studies have focused almost exclusively on global outcome measures (Durmer & Dinges, 2005; Ratcliff & van Dongen, 2009; van Dongen, 2004). Finally, compared with the number of published studies on the effects of acute total SD (>45 hours and 17-45 hours), there are far fewer studies detailing the effects of chronic partial sleep deprivation of reduced sleep (eg. <7 hours/ 24 hours) over many days, weeks or months. The reason for this is most likely due to the fact that such longer term studies are expensive to run and more difficult to control in terms of both environmental and individual factors (eg. not permitting use of any stimulants, requiring greater participation time). What this does mean however, is that there are prominent gaps in the literature around the characterisation of performance and behaviour following chronic sleep loss. For the applied field setting, chronic partial SD is much more likely to be a representative form of sleep loss in real world operations and modern society in general, especially when taking into account various medical conditions, sleep disorders, work demands, and social and domestic responsibilities (Banks & Dinges, 2007).

The research findings on the effects of sleep loss on the full myriad of cognitive abilities is beyond the scope of this review to cover in full. However, prominent findings generated by reviews of the literature that are of greatest relevance to safe pilot performance and safety performance standards expected of the pilot are discussed below, and a practical synthesis of the findings informed by extensive reviews of the literature is provided Table 1 (Durmer & Dinges, 2007; Ahola & Polo-Kantola 2007; Goel et al., 2009).

Table 1. Summary of Neurobehavioural and Cognitive Performance Effects of Sleep Deprivation

Involuntary microsleeps occur
Attention-intensive performance is unstable with increased errors of commission and omission
Cognitive slowing occurs in subject-paced tasks, while time pressure increases cognitive errors
Response time slows
Both short term recall and working memory performance declines
Reduced learning (acquisition) of cognitive tasks
Performance requiring divergent thinking deteriorates
Response suppression errors increase in tasks primarily subserved by prefrontal cortex
Response perseveration on ineffective solutions is more likely
Increased compensatory effort is required to remain behaviourally effective
Tasks may be begin well, but performance deteriorates as task duration increases
There is growing neglect of activities judged to be nonessential

Aside from its impact on alertness, sleep loss exerts the largest detectable performance effects on attentional processes. Whilst at a general level, sleep loss has been found to have a detrimental impact on multiple aspects of cognition, cognitive tasks vary considerably in their sensitivity to sleep loss (Waters & Bucks, 2011). The effect sizes of most cognitive deficits as a result of sleep deprivation are in the 'moderate range', with the largest effect sizes on tasks which assay processing speed and vigilant attention (Lim & Dinges, 2010). Indeed, laboratory studies have provided strong empirical support for the finding that subjects' reaction time, processing speed and vigilant attention show the greatest deficits as a result of SD, compared with smaller, more inconsistent decrements in cognitive faculties such as short term memory, mental arithmetic, and other higher executive functions (Durmer & Dinges, 2005). For example, performance on Baddeley's Logical Reasoning Test (a fairly complex task) is consistently found to be stable, even as sleepiness increases and decrements on simple cognitive tasks (such as Reaction Time or PVT) appear (Magill et al., 2003; Smith & Maben, 1993). The psychomotor vigilance test (PVT; Dinges & Powell, 1985) is a 10 minute simple reaction time task that requires the subject to respond as quickly as possible to a visual cue that occurs at a pseudo-random interval ranging between 2 -10 seconds. The PVT is sensitive to the slowing of responses and increased attentional lapses, and shows very little effect of learning, making it a useful tool for repeated administrations over the course of an experiment (Doran et al., 2001). Hence most investigations into the impact of sleep loss on vigilant attention have used this measure. A meta-analysis by Lim and Dinges (2008) set out a number of conclusions from SD studies using this test. These authors note that typically sleep loss impacts attentional performance on tasks such as the PVT by a general slowing of response times, both with respect to the average reaction time for trials and the fastest 10% of responses. In addition, sleep loss produces increased errors of omission and general variability in performance on the PVT, in both acute and chronic forms of SD (van Dongen et al., 2003). Lim and colleagues further outline the finding that SD leads to an increase in the number and duration of lapses (response times exceeding 500ms) as well as an increase in errors of commission (responses where cues are not present). The progressive deterioration of performance with increasing time spent on the task (the time on task effect) evidenced in the PVT and many other attentional tasks has been found to be greatly exacerbated when subjects are sleep deprived (Lim & Dinges, 2008). Finally various studies have shown that the PVT is highly sensitive to both homeostatic and circadian drives (Basner et al., 2011; Doran et al., 2001; Goel et al., 2015), with the greatest deterioration in subjects' performance typically occurring during the early morning hours. As vigilant attention appears to be so consistently and robustly affected by sleep deprivation, some authors have suggested that deficits in cognition can be explained by the central function of sustained attention in many higher order cognitive tasks (Balkin et al., 2008; Durmer & Dinges, 2005; Lim & Dinges, 2008, 2010).

Sleep loss is also associated with detrimental effects on learning and working memory. Indeed, separate lines of research have suggested sleep loss adversely affects verbal learning, memory acquisition and retention, (Alhola & Polo-Kantola, 2007; Goel et al., 2009; Walker & Stickgold, 2006). As prefaced earlier, findings from different studies on aspects of higher cognition are somewhat difficult to generalise, given prominent differences between experimental tasks used and the focus of most efforts solely on acute forms of sleep deprivation. Nevertheless, Lim and Dinges (2010) in summarising their meta-analysis findings on working memory concluded that sleep deprivation does adversely affect both accuracy and response times of memory tasks, with effect sizes generally in the moderate range. Most efforts have focused on tasks that assess learning or encoding of new information, or tests that attempt to assess the consolidation and assimilation of newly learned information into existing memory. In the former category, cognitive tests where decrements of performance have been reported during acute TSD include n-back style and choice-reaction time tasks (which carry a working memory component) (Choo et al., 2005; Jennings et al., 2003; Smith et al., 2002), verbal memory tasks (Chee et al., 2006; Mu et al., 2005) and types of single unit recall (Frey et al., 2004). However, it should be noted that other acute TSD studies on working memory recall, and visuo-spatial memory have not yielded statistically significant performance decrements (Heuer et al., 2005; Nilson et al., 2005; Quigley et al., 2000). More recently it has been suggested some components of memory processes are more vulnerable to the effects of sleep deprivation that others. For example, one night of TSD appears to impact visual working memory in terms of ‘filtering efficiency’ (via the ability to ignore distracting stimuli), but not capacity (Drummond et al., 2012). Research initiatives using neuroimaging techniques have helped provide some insights into the variability of these performance outcomes and mixture of research findings with respect to sleep loss and memory. Indeed, these have suggested that sleep deprivation has a detrimental impact on normal hippocampal and medial temporal lobe functioning, thus hindering their primary function of facilitating the formation of new memories. However, where this happens, other brain areas can act to facilitate memory performance. For example. Drummond and colleagues found that 35 hours of sleep deprivation resulted in significant decrements in verbal learning compared with rested wakefulness (Drummond et al., 2000). Neuroimaging analysis showed significantly reduced activation within the medial temporal lobe regions (including the hippocampus) in sleep deprived subjects, but increased ‘compensatory’ activation within prefrontal and parietal cortices. Upon analysis of the subsequent performance findings, increased activation of the prefrontal and parietal areas was associated with better performance on the recall test during sleep deprivation. Thus, it seems that on some memory tasks, other brain areas may be able to take over some of the functioning to help preserve performance under conditions of SD. However, where other brain areas are not able to be recruited, performance is thought to suffer from greater deterioration. Supporting this hypothesis, Mu et al., (2005) showed that impairment in working memory performance under 30 hours of SD was associated with corresponding reductions in prefrontal and parietal activation levels. Thus performance outcomes on various learning and memory tasks may substantially relate to the degree to which other compensatory processes (particularly in the prefrontal or parietal brain areas) can be recruited to preserve performance in the individual participant, and this may underpin some of the individual differences in sleep-driven performance decline.

Some authors have gone further to suggest that sleep loss can have a considerable impact on the prefrontal cortex (PFC), which means that tasks that recruit this brain region may be particularly vulnerable to performance decrements (Babcoff et al., 2005; Beebe & Gozal, 2002; Curcio et al., 2006; Harrison & Horne, 2000). Whilst the PFC may be involved in a number of aspects of cognition, functional magnetic resonance imaging (fMRI) studies of divergent thinking as well as tasks that involve working memory and executive function (manipulation and maintenance of information in one's 'working' attention and the minimization of distracting or irrelevant material) have most consistently activated the ventral and dorsal areas of the PFC (Fink et al., 2009; Goel & Vartanian, 2005; Kröger et al., 2012). In response to sleep loss, the PFC shows elevated activation levels in relation to verbal learning and reasoning (Drummond et al., 2000, 2004, 2005; Jonelis et al., 2012), which may be interpreted as a compensatory function of the brain to counteract the SD impairment, in the form of increased activity. However, some studies investigating working memory tasks have shown that SD can lead to decreased PFC activity (Chee & Chuah, 2008), suggesting that increases or decreases in PFC activity may actually vary depending on task difficulty, with more difficult tasks being more likely to elicit a compensatory response (Drummond et al., 2004; Vartanian et al., 2014).

Some research has suggested that performance on 'crystallized' abilities and learnt schemas are the least affected of all cognitive domains on the basis of available evidence. Indeed, a wide-ranging literature review conducted by Harrison and Horne (2000) on decision making suggested that the complex cognitive processes engaged in tasks assessing broad intellectual functioning, reading comprehension, logical deduction and critical reasoning does not significantly degrade performance, even when subjects undergo two nights of TSD. Supporting this finding, Lim and Dinges (2010) in their meta-analysis found no significant effect of sleep deprivation on tests of crystallised intelligence, reasoning and various types of problem solving. With regards to problem solving however, it is important to note the differential effect of sleep loss on convergent thinking and divergent thinking. In contrast to the findings above, tasks that require novel or divergent thinking (ie. testing fluency, originality, elaboration, and flexibility) appear to be significantly affected by sleep loss

(Harrison & Horne, 2000). Although this domain of cognition is more challenging to test, evidence from an early study by Horne (1988) showed that subjects following 32 hours of acute SD provided less creative responses, fewer unusual or original ideas, reduced ability to shift thinking strategy and a reduced ability to generate novel words to phonemic cues than well rested controls. Horne (1988) in explaining these findings suggested that this effect was not accounted for by loss of motivation in participants, but rather that "sleep loss made them fixate on previously successful strategies when attempting solutions to the next problem" (Horne, 1988, p535). The same study furthermore showed that sleep deprived subjects required longer planning time and showed greater perseverations on the Tower of London Test, a task designed to assess planning and ability to shift strategy. More recently, Killgore et al. (2009) found that two nights of acute SD increased the number of moves taken to solve the Tower of London Task. Interestingly, administration of caffeine (which given similar homeostatic and circadian drive pressures re-establishes baseline levels on PVT tasks) was not seen to improve performance in these sleep deprived subjects (Killgore et al., 2009). Wimmer et al. (1992) furthermore demonstrated that divergent thinking performance (in a task assessing subjects' conceptual diversity of generated solutions) was degraded in subjects following just a single night of SD. In explaining these differences some authors have suggested that convergent thinking tasks rely substantially less on prefrontal areas, and so performance on these deductive reasoning tasks, or tasks that rely on established knowledge is better preserved under conditions of SD. By contrast, tasks requiring divergent thinking or innovative problem solving may critically rely on PFC functioning (Vartanian et al., 2014). From an operational perspective, these findings suggest that pilots may be less vulnerable to the effects of sleep loss when conducting certain procedures that have been 'embedded' from their training, or whilst undertaking activities such as standard operating procedures, which follow a logical sequence. By contrast, safe performance and decision making during novel operational scenarios which require divergent thinking may be an aspect of performance particularly threatened in sleep deprived pilots.

Early research on SD and cognition suggested that sleep loss generally reduces the motivation of participants to perform on tasks (Wilkinson, 1961), leading certain authors to suggest that both novelty and motivation are critical factors affecting task performance under SD (Wilkinson, 1961; Williams et al., 1959). This hypothesis, which later prompted the formation of the 'controlled attention model' (Pilcher et al., 2007) came from the initially surprising observation that following short periods of total SD, performance on many highly demanding cognitive tasks remains intact, whilst on simple cognitive tasks performance degrades as extended wakefulness accumulates. In light of these findings, proponents of the Controlled attention model, made a similar prediction that performance is more severely affected by SD on monotonous, or intrinsically less engaging tasks, due to the need for greater 'top-down' control to maintain optimal performance. Conceptualised in this way, the authors do not limit their explanation of human performance exclusively on intrinsic task or environmental characteristics (Pilcher et al., 2007, 2013), but rather consider the participants' ability to control their own attention and engagement in a given task as a central factor driving the differences in impact of SD on performance. Interestingly, while other theorists have offered different explanations for the same findings, arguing instead that arousal and vigilance are general factors that explain much of the variance in the degree cognitive deficits accrue following sleep loss (Durmer & Dinges, 2005; Lim & Dinges, 2008), both accounts make similar predictions concerning cognitive performance loss with respect to simple (vigilance based) cognition verses more complex forms of higher cognition. The controlled attention theory proponents however suggest further that tasks that intrinsically encourage attentive behaviour and are subjectively varied or interesting to the individual will be the least affected by SD. Harrison and Horne's (2000) review presents some studies that challenge aspects of this hypothesis, as sleep deficits are found on high novelty 'interesting' tasks, arguing against the idea that intrinsic task interest is necessarily a critical contributor to performance under SD conditions (Harrison & Horne, 2000). However the same studies do not appear to attempt to assess subjective interest of individual subjects in the tasks at hand, which may still play a particularly important role in performance.

Although historically there has also been less focus on the effect of SD on important influencers to performance such as mood state, existing studies suggest that practically all forms of sleep deprivation appear to cause a general increase in negative mood states, including feelings of weariness, irritability, fatigue, loss of vigor and confusion. As Kilgore (Killgore, 2010) points out, any person who has gone without sleep for a night is likely to have experienced a decline in mood and a rapid increase in irritability and emotional volatility. Sleep deprivation and circadian disruption have been shown to increase the probability of negative emotions (Baglioni et al., 2010; Caldwell et al., 2004; Short & Louca, 2015), with some research suggesting that acute SD for just one night produces a significant increase in negative self-rated mood scores, compared with rested controls (Tempesta et al., 2010). Beyond these broader findings, critiques of research initiatives investigating the effects of sleep loss on mood suggest that many early studies may lack requisite fidelity and not control for confounding factors such as experimental environment. For example, an early meta-analysis (Pilcher & Huffcutt, 1996) suggested that the detrimental effects of sleep deprivation on feelings of fatigue and related negative mood states were greater than those evidenced in cognitive performance tasks. However, this conclusion might have been based on inadequate experimental controls and cognitive assessments in early partial SD studies (Goel, 2009). Indeed, more recent research initiatives have suggested that chronic SD produced much more rapid cumulative detriments in cognitive performance, than associated subjective measure of fatigue (Van Dongen et al., 2003; Banks & Dinges, 2007; Belenky et al., 2003).

It should be noted that sleep-driven cognitive decline is thought to be both clinically normal and typically reversible with a return to normal sleep. However, the length of sleep and overall time required to recover to normal cognitive performance levels, and the processes underpinning this recovery are less well established. In a number of sleep deprivation studies, the subsequent return to normal baseline performance is not assessed or reported. As highlighted earlier, recovery sleep does appear to show more rapid sleep onset and transition to SWS and REM-sleep, than normal sleep, and the relative proportions of different phases of these types of sleep may depend on different factors such as whether sleep loss has been acute or chronic (Alhola & Polo-Kantola, 2007). Indeed, one sleep period (of at least eight hours) appears to reverse the adverse effects of acute SD on simple forms of cognition, such as performance on the PVT (Brendel et al., 1990; Corsi-Cabrera et al., 2003; Drummond et al., 2006; Kendall et al., 2006). By contrast, studies on chronic partial SD have found that PVT performance is not restored so rapidly. Dinges et al. (1997) in a study with seven consecutive sleep restriction nights of 5 hours per night found that performance only approached baseline levels following two 10 hour nights sleep. Similar findings have been reported by Belenky and colleagues (Belenkey et al., 2003). However, two separate fields of evidence suggest more research on recovery rates with different aspects of cognition is necessary. First, a study using positron emission tomography (PET) neuroimaging has shown that following 24 hours of SD, one night of recovery only partially reversed metabolic reductions in the thalamus, basal ganglia, and frontal lobe regions (Wu et al., 2006). Secondly, van Dongen and others have pointed out that there may be considerable inter-individual differences in recovery rates, which may act to confound broad study conclusions on recovery rates, where these differences are not taken into account (Van Dongen et al., 2004).

In summary, sleep deprivation causes deterioration in performance on a wide variety of cognitive functions. However, cognitive tasks and individuals vary considerably in their sensitivity to sleep loss. In general, although particularly for simple attention or vigilance-based cognitive tasks, performance becomes progressively worse when time on task and continual wakefulness is extended. This primary evidence base is important to consider for understanding which aspects of operational performance are at greatest risk of decline with increasing levels of operator sleepiness. Most of the insights concerning the effects of SD on performance have been gleaned from acute TSD laboratory based experiments, which may limit some of their operational applicability. However, studies assessing the performance decrements associated with continual wakefulness are important to reference with respect to extended shift hours and roster patterns that drive sleep loss. Moreover, evidence of the wide-ranging impacts of sleep loss on brain functioning (lapses, involuntary sleep, declines in visual awareness), cognitive performance (impaired vision and response times, components of working memory and learning), and mood (irritability, negative emotion) are important to consider when comparing symptoms and rates of ‘pilot fatigue’ with existing acceptable standards of performance for the pilot in commercial aviation.

Sleep deprivation and driving performance

In the applied field of transport research, the most extensive experimental investigations into sleep-related performance decline have been with respect to driver fatigue and road safety, as sleepiness or falling asleep while driving has historically accounted for a considerable proportion of all vehicle accidents (Horne & Reyner, 1999; Shekari Soleimanloo et al., 2017), with some estimates world wide of around 20% of motor vehicle accidents (MacLean et al., 2003). Indeed, sleepiness in night shift workers is a major risk factor for vehicle accidents (Horne & Reyner, 1995; Lyznicki et al., 1998; Philip et al., 2005), as drivers show increased micro-sleep propensity, periods of inattention, more numerous and lengthy eye closures, and an increased likelihood of falling asleep at the wheel (Åkerstedt et al., 2005; Lyznicki et al., 1998). High fidelity simulator studies of drivers have suggested that increasing levels of sleepiness are associated with general performance deficits in lane deviations and difficulty maintaining lateral positioning (Ingre, Åkerstedt, Peters, Anund, & Kecklund, 2006; Otmani et al., 2005), increased heading errors, reduced steering activity, and reduced perceptual sensitivity (Matthews & Desmond, 2002), greater variability in speed and slow reaction time to ‘on-road’ events (Arnedt et al., 2000; Lenné et al., 1998; Philip, Sagaspe, Moore, et al., 2005). In a systematic review of driver performance in controlled experimental settings it was concluded that in line with sleep science predictions, objective and subjective measures of sleepiness increase with driving duration (Liu et al., 2009). Although most studies carry the limitation of not investigating individual differences, of particular interest is that several studies have suggested that ‘drowsy driving’ performance can be largely characterised by increased probability of lane departures through reductions in small steering wheel movements (to counteract drifting across lanes) when drivers lapse or lose concentration, followed by increased large steering movements towards the lane centre (to correct normal drifting deviations) (Liu et al., 2009). These findings hence offer useful applied field context, as they suggest that at least for some transport settings, there may be hallmark manifestations of fatigue-related operator performance decline.

1.6. Sleep deprivation and pilot performance

As noted by various authors, the role of modern commercial pilots has substantially changed over the past 40 years, with modern aircraft largely being coordinated by the pilot via inputs to a computerised system that communicates to the control surfaces and engines (Endsley et al., 1998). Importantly the inclusion of an autopilot system as part of the flight control system means that pilots now spend much less time during a flight manually controlling the aircraft (which is typically reserved for the critical phases of flight during take off and landing). As such, there is a much greater role for pilots in the monitoring of the flight deck systems and indicators, alongside the management of communications and any external or internal threats that may manifest during flight (Roach et al., 2012). However, piloting a modern commercial jet aircraft is still considered a complex and demanding activity, that requires highly specialized skills, discipline in maintaining standard operating procedures and perhaps most importantly, problem-solving in the presence of uncertainty and risk (Endsley, Todd, et al., 1998; O’Hagan et al., 2020). In emergency situations, this may require rapid but prudent decision-making based on knowledge of the aircraft, environmental factors, accompanying flight crew and one’s own abilities (Kern, 1997). Moreover, pilots may have to continuously integrate inputs from external and internal data sources alongside past experience in order to understand an evolving in-flight situation that may be entirely novel. In operational circles this is sometimes referred to as having ‘situation awareness’, where the pilot’s mental model of the situation is updated along with in-flight developments and relied upon for their judgements and decision making (Endsley, Farley, et al., 1998). In normal operations, the vulnerability of the aircraft automated system design with respect to pilot performance is that in abnormal conditions and low frequency events, which represent the highest workload and greatest risk to safe flight, pilots are expected to quickly notice signs of deviations, and actively take over control (Bainbridge, 1983). Thus it is not only vigilant attention and alertness that are pre-requiste cognitive functions required for normal pilot performance, but also a myriad of higher cognitive functions that are relied upon to maintain safe flight.

Survey and self-report findings

Survey and field studies conducted with pilots suggest that inflight drowsiness, microsleeping and related involuntary sleep phenomena have been experienced by the majority of pilots during their duties. In 2013, a representative sample of 500 pilot members of the British Airline Pilots’ Association were asked via the COMRES polling agency about involuntary sleep on the flight deck. Over half (56%) of the respondents admitted that they had involuntarily fallen asleep during two-crew operations, and of these individuals, almost a third (29%) admitted to having had at least one occasion where they had woken up to find the other pilot also asleep (ComRes, 2013). Similarly, Rosekind and colleagues found that nearly three-quarters (71%) of pilots reported that they had ‘nodded off during a flight’ in a study of nearly 1172 pilots (Rosekind et al., 2000). Whilst not direct measures of involuntary sleep, other studies with commercial pilots have consistently highlighted that pilots report increasing subjective sleepiness and pressure to sleep during their duty periods (Gundel et al., 1995; Petrie et al., 2004; Petrilli et al., 2006; Roach et al., 2010), and may continue to operate despite feeling severely fatigued (Jackson & Earl, 2006). Some studies have also indicated that high proportions of pilot suffer from clinically significant levels of fatigue, as determined by elevated fatigue severity scores (FSS) of 5 in 45% of British Pilots (Steptoe & Bostock, 2011) and FSS 4 in ± 90% of Portuguese pilots (Reis et al., 2013, 2016).

Other sources of empirical evidence offer support to these self-report findings and suggest that the true incidence of elevated levels of sleepiness and involuntary sleep events experienced by flight crew may in fact be higher than the level that could technically ever be captured by subjective report methods. For example, field studies using EEG and EOG methodologies and other physiological parameters have also suggested that significant drowsiness and involuntary sleep events occur on long-haul flights. One research initiative conducted on long-haul commercial flights with a UK registered airline between London and Miami showed that 10 out of the 12 pilots included in the study either slept or showed evidence of sleepiness during their flight, as assessed by quantitative analysis of the EEG and EOG recordings (Wright & McGown, 2001). Across this group of pilots, 38.1 minutes were characterised by isolated incidents of sleepiness (i.e. not associated with actual sleep) during the night and 4.2 minutes were evidenced during the daytime. The majority of these lapses in wakefulness lasted less than 20 seconds, and the authors suspected that the pilots may well have been unaware of some of these occasions. Using EEG recordings, involuntary sleeps on the flight deck lasting longer than 10 minutes (Graeber et al., 1990) as well as periods of simultaneous sleepiness in both captain and co-pilot (Cabon et al., 1993)have furthermore been reported, and other studies have identified slow eye movements, microsleep and sleep events in pilots during flight (Rosekind et al., 1994; Samel et al., 1997). As described earlier in this review, such bouts of involuntary sleep or sleepiness-related lapses appear to be inevitable neurobehavioral consequences of the destabilisation of the waking state in humans (Lim & Dinges, 2008).

When considering themselves, pilots report effects of fatigue on in-flight performance which are consistent with established scientific findings on the main cognitive effects of sleepiness (Gregory et al., 2010; Rosekind et al., 2000). These include degraded alertness, slowed reaction time, inability to concentrate, errors of omission and commission, deterioration in judgement and decision making, worsened mood, and deteriorating flying skills (Bourgeois-Bougrine et al., 2003; Gregory et al., 2010; Rosekind et al., 2000). When pilots become more tired, they report that concentration during supervisory and monitoring activities becomes more difficult than usual (Bourgeois-Bougrine et al., 2003) in line with laboratory findings from simple vigilance tasks. Moreover, pilots report that fatigue on duty causes more small mistakes (in terms of calculation or interpretation) and compromised communication (Bourgeois-Bougrine et al., 2003). Other authors have highlighted the fact that affective changes following sleep loss may be particularly important for safety, as they may lead pilots to take unnecessary risks or become overly stressed or frustrated (Drury et al., 2012; Hartzler, 2014; Neri et al., 1992).

Performance consequences of sleep deprivation in simulated flight

Research initiatives using high fidelity flight simulators reveal similar findings to those evidenced in simulated driving with respect to the impact of sleep loss on symptoms and cognitive performance. However, there are very few studies of this kind, and most have been conducted with military pilots in military aircraft simulators, where the remit of flying tasks and flying environment notably differs to that of commercial aviation. Setting these considerations aside, studies with sleep-deprived military pilots suggests that pilots' control of even the most basic flight parameters declines significantly following 18 to 24 hours of constant wakefulness (Caldwell et al., 2009; Previc et al., 2009). In one study of F-117 pilots (Caldwell et al., 2004), following one night of sleep loss, there was a twofold increase in control errors (e.g., airspeed and altitude deviations) on precision instrument manoeuvres (e.g., straight and level flight, climbs, and descents). These decrements in performance were notably accompanied by changes in mood, affecting crew communication. Following 24 hours of continuous wakefulness, levels of confusion and fatigue increased, and there were substantial elevations in slow-wave EEG activity (of the type usually associated with extreme drowsiness). Thus extended wakefulness in aircrew can produce decrements in performance that can ultimately lead to significant operational problems. Indeed, this point was illustrated by Bartlett (1953), who in an early experimental study on aircrew monitoring found that increasing levels of subjectively rated 'fatigue' was associated with increasingly larger deviations of instrument readings, as well as elevated levels of distraction, cue restriction (and neglecting peripheral cues), and response variability. More recently, one study found that US Air Force pilots demonstrated increased visual neglect for stimuli both in the central and peripheral visual fields of awareness during a simulated 12-h overnight flight (Russo et al., 2004). Some authors have also suggested that with increasing levels of sleepiness, task-related details associated with safe flight are often overlooked or forgotten, and pilots lose the ability to integrate information from individual flight instruments due to 'attentional narrowing' (Caldwell & Caldwell, 2003). In sum, the effects of sleep deprivation have been found to impair the key aspects of pilots' duties and 'traditional' piloting skills: their ability to attend to flight instruments, make correct decisions, communicate on the radio and with crew and to perform well on navigational tasks (Caldwell, 2012; O'Hagan et al., 2020; Wilson et al., 2006).

Although the research initiatives with military pilots provide many valuable insights on the nature and types of performance decrements associated with sleepiness on the flight-deck; both the operational environments of commercial jets, and the nature of the flight task demands differ substantially to those of the military operations. Increased cockpit automation has fundamentally changed the tasks that commercial pilots today face, and likely plays a key role in increasing the symptoms of fatigue and lapses in vigilance experienced on the flight-deck. Indeed, as Hartzler (2014) points out, within the cruise phase of long-haul flights, pilots are likely to suffer the combined effects of monotonous environment, habituation, deteriorations in vigilance and increased boredom which may increase subjective sleepiness. In addition, pilots must regularly contend with other environmental factors that can impact alertness levels such as dim lighting on the flight deck (Graeber et al., 1990) and the absence of tactile feedback from the controls (Roach et al., 2012). In daily commercial flights, manual handling has largely been substituted with systems monitoring. Increasing levels of automation have furthermore helped to reduce the objective workload of the flight crew (Hartzler, 2014), potentially acting to increase task-related monotony for the cruise phases of flight, and in turn, on-duty sleepiness. Thus some caution should be taken in extrapolating the findings of simulator studies conducted with military pilots to the commercial aviation setting.

One of the only large-scale simulation studies (n=67) to directly investigate the relationship between prior sleep-history and piloting performance in a modern commercial jet type (Boeing 747-400) was conducted by Thomas and colleagues (2006; described in Gander & Signal 2008). The simulation scenario was described as containing a series of realistic threats and a critical decision event about whether to divert the flight to a different destination. Here the authors reported that pilots who had both obtained less than 5 hours sleep in the last 24 hours committed 14.3% more errors in response to a series of realistic in-flight scenario threats compared with crews that had achieved more than 5 hours sleep. However, perhaps due to compensatory behaviours, interestingly the sleep restricted crews were more likely to detect these errors (59% detection, compared with those 48% detection). Using multiple logistic regression analyses, the authors further found that the sleep loss of the captain played a significant role in the safe performance of the multi-crew element. For every 1 hour less sleep obtained by the captain, the likelihood of an error being mismanaged was reported to be increased by 40% (Gander & Signal, 2008; Thomas et al., 2006). To the best of knowledge, the results of this study were only presented at an industry seminar for peer review, and as such have not been published in their own right. However, as one of the only studies of its kind, the findings offer some promising evidentiary support for the value in investigating sleep-wake history parameters within commercial simulator based operations. Moreover, these initial findings suggest that even more partial forms of sleep restriction can act to deteriorate commercial pilot flight performance in a multi-crew setting and raise interesting questions regarding the relationship between increased sleepiness in flight crew and safety-related behaviours. Thus, a key gap in the aviation literature relates to the dose dependent effect of sleep deprivation on flying task performance in the commercial jet environment. Given the relative success of driving simulator studies in identifying common characteristics of fatigue-related driver performance decline, future work would benefit from more simulator studies of this type for commercial aviation.

1.7 Safety consequences of pilot fatigue

As explained in the previous section, there is limited simulator evidence with respect to sleep loss and its impact on commercial flight performance. As such, further research is needed to understand the relationships both between increasing pilot sleepiness levels and flight performance, and the resultant impact of sleep-driven performance decrements in terms of the overall safety of flight. Part of the reason why this latter relationship has been so difficult to establish is that commercial aviation systems have been developed to high reliability standards. For the last 30 years, commercial air transportation, has maintained its record of being one of the fastest and safest modes of transportation (Roach et al., 2012). As such, there are relatively few data regarding the causal factors associated with commercial aviation accidents in general, compared with other transport industries. One approach to appraise this latter relationship is to try to compare the fatigue hazard of pilot performance with other accepted human performance standards (discussed further in Section 1.8). However, in addition, it is helpful to appraise what available accident or incident data there are relating to sleep and circadian drivers and ‘pilot fatigue’ and to consider survey data from pilots on their perceptions of the safety hazard of fatigue.

Fatigue in aviation incidents and accidents

Pilot fatigue has been implicated as either a causal or contributory factor to aircraft crashes and serious incidents on a number of occasions (Marcus & Rosekind, 2017; National

Transportation Safety Board (NTSB), 1994, 2010, 2017; Swiss Aircraft Accident Investigation Bureau, 2001). Where pilot fatigue is implicated, investigators have typically used evidence-based inferences from sleep science principles, to determine whether 'fatiguing' circumstances are likely from the pilots' rostered duty hours or other forms of evidence that indicate the flight crews' likely sleep-wake history preceding the duty where the accident occurred. For example, two aviation accidents that clearly highlighted a role of severe acute sleep deprivation as a causal factor include the 1993 American International Airways Flight 808 crash, and the 2009 crash of Continental Connection Flight 3407. Both flights were on approach to destination airports when the aircraft ended up stalling and crashing while attempting to land. For Flight AA808, it was discovered that the captain had been awake for 23.5 hours and had slept for 5 of the 28.5 hours prior to the accident (NTSB, 1994). For Continental Connection Flight 3407, accident investigators determined that 'the pilots' performance was likely impaired because of fatigue' because before flight, one of the two flight crew had been awake all night, and the other had reported for duty following a lengthy commute and a nonrestorative sleep period (NTSB, 2010). Both accidents hence provided evidence of acute fatigue risks at the pilot level, arising from extended hours of wake and sleep deprivation prior to flight.

An accident where investigators explicitly referenced an extended period of prior sleep and duty considerations as well as the duty time came from the Swiss Aircraft Accident Investigation Bureau in their documentation of the 2001 Flight CRX 3597 crash. Here the pilots of Flight CRX 3597 collided with treetops in conditions of low visibility on their approach into Zurich airport, before subsequently crashing into the ground, where the airframe caught fire on impact. The Swiss Aircraft Accident Investigation Bureau in their final report (Swiss Aircraft Accident Investigation Bureau, 2001) found that on the two days prior to the accident, the captain had exceeded maximum duty times and had reduced rest time the night before the accident. On the day of the accident, the commander had been awake for at least 15 hours and on duty for more than thirteen and a half hours. The authors concluded that, as a contributing factor to the accident, 'fatigue [had] negatively affected the pilot's ability to concentrate and make appropriate decisions and impaired his capacity to analyse complex processes'. As recently as 2017, an exceptionally close call involving Air Canada flight 759 highlighted the vulnerability of overreliance on regulatory controls of hours for management of fatigue risks. In this case, reserve pilots approached San Francisco airport at 23.56 local time (roughly 02:56 for the flight crews' normal body clock time), with the captain having been over 19 hours awake due to being called from standby late within the allotted period (NTSB, 2017). Misinterpreting the taxiway for the right-hand runway, the pilots only narrowly avoided colliding the aircraft with four grounded aircraft, all fully loaded with passengers, waiting on the taxiway. In this serious incident, the NTSB concluded that the fatigue likely played a role in the crewmembers' degraded perception and that the event highlighted the fact that the applicable Canadian regulations controlling pilot hours 'do not, in some circumstances, allow for sufficient rest for reserve pilots' (NTSB, 2017, p70). Taken together, these accidents and serious incident examples highlight how both duty and prior sleep and wake factors may contribute to increased in-flight fatigue-related risks that contribute to unsafe pilot performance. In addition, three of the four examples provided also highlight the systems level vulnerability of both flight crew being simultaneously affected by fatigue-related performance decline or incapacitation.

Reviews of commercial aviation crash reports have concluded that in at least 4-8% of crashes, fatigue is likely to have played a contributory role (Caldwell, 2005), and that duty time is linked with an increased likelihood of crash risk (Goode, 2003). However, some caution should be taken over the broader extrapolation of these rates, since the processes by which crashes or serious incidents have previously been categorised as 'fatigue-related' are not entirely known and may depend on available circumstantial evidence (Lyman & Orlady, 1981; Pouliquen et al., 2005). Aviation accident investigators historically have not had a systematic approach for assessing the potential influence of operator fatigue in accidents and serious incidents, nor included an assessment of duty schedules as part of their routine investigations[2]. Whilst the basic trajectory of the aircraft, and inputs into the flight computer by pilots are monitored in scrupulous detail within accident report engineering analyses, the 'fatigue status' of the human operator during normal flights is not routinely measured or monitored, and available post-hoc data pertaining to pilot fatigue may be circumstantial. For example, in the 2010 crash of Air India Express Flight 812 into Mangalore, the cockpit voice recorder indicated that the captain had been asleep for the first 1 hour and 40 minutes of the flight (Court of Inquiry India, 2010). According to the NTSB, this was one of the first accidents where snoring had been recorded on a cockpit voice recorder shortly before an incident (Wingelaar-Jagt et al., 2021). Historical definition debates on what is meant by 'fatigue' and a lack of standard measurement index (Brown, 1994) have also meant that the

---

[2] Personal communication with investigators from the AAIB (2016)

consideration of fatiguing factors or the process by which occurrences are categorised as 'fatigue-related' by most accident investigators may differ across the world. For example, Lyman and Orlady (1981) found that 'fatigue' was implicated in 77 (3.8%) of 2,006 incidents reported by pilots to the Aviation Safety Reporting System (ASRS), a voluntary safety occurrence database in America. However, when their analyses were extended to include all factors that could be directly or indirectly linked to fatigue, incidents potentially related to fatigue increased to 426 (21.2%) (Lyman & Orlady, 1981; Rosekind et al., 1995). In a more recent study, fatigue was estimated to have been a probable cause, contributing factor or finding in 23% of aviation safety incidents or accidents investigated by the NTSB between 2001 and 2012 (Marcus & Rosekind, 2017). Thus there are varying estimates over the proportion of safety events where fatigue has been identified as a contributory factor, but more recent estimates have been towards 20%. Concerns over the accuracy of the role of fatigue in incidents and safety occurrences have also been shared by some aviation safety regulators. For example, in 2006, the New Zealand Civil Aviation Authority (NZCAA) reported that 0.2% of incidents had fatigue as a contributing factor while in Australia the level was reportedly 7%. However, both the Australian Transport Safety Bureau (ATSB) and the NZCAA have suggested that such estimates are 'unrealistically low' and instead estimate that the real rate is closer to 25% (Signal et al., 2006). Thus, accurate information on the contribution of fatigue to aviation accidents or incidents is often difficult to obtain, and rates may vary considerably depending on the reporting system, the timing and nature of questions asked, and the skill and knowledge of investigators (Signal et al., 2006). As a consequence, coding issues are likely to limit meaningful and systematic extrapolation of fatigue risk rates from accident and incident historical databases.

Workforce perceptions of pilot fatigue as a safety hazard

Concerns over the safety impact of pilot fatigue are widespread across the national and international pilot workforce. Pilots frequently link their experience of fatigue with unsafe piloting performance in the cockpit (Bourgeois-Bougrine et al., 2003; Reis et al., 2013; Rosekind et al., 2000; Rudari et al., 2016), and have attributed incidents, errors or operational problems due to fatigue in one-third of all reports submitted to the UK Confidential Human Factors Incident Reporting Programme (CHIRP) (Jackson & Earl, 2006). In the UK, CHIRP findings may offer a closer approximation of the burden of fatigue during normal operations than other industry safety databases, since pilots frequently underreport their fatigue levels via company and regulatory channels. Indeed, by way of example, Reis et al. (2013) found that 91.4% of pilots sampled claimed to have ‘made mistakes in the cockpit as a direct consequence of fatigue’, yet 81% had not *ever* reported themselves unfit for flight due to fatigue. Where British pilots have been surveyed over the past decade, sleepiness (or ‘fatigue’) has been consistently highlighted by respondents as the greatest threat to flight safety, up to three times more than any other threat, and 40-45% report that their flying abilities are compromised by fatigue at least once a month (ComRes, 2013). Similarly, Rosekind et al., (2000) reported that 85% of American pilot survey respondents identified fatigue as a ‘moderate’ or ‘serious’ safety issue. When questioned on the extent to which fatigue affects the safety of flight operations, 93% of German pilots, and over 70% of pilots from Sweden, Norway and Denmark report having made mistakes whilst flying due to fatigue (European Cockpit Association, 2012).

1.8 Safety standards and the human component within aviation

Despite widespread recognition that sleepiness degrades neurobehavioural performance in humans, and further operational evidence implicating ‘pilot fatigue’ as a risk factor in commercial flights, it has been difficult for practitioners to determine at which point this risk becomes unacceptable. Data from other transport and operational settings suggests that there are elevated fatigue-related risks where continual wakefulness exceeds 16 hours, the sleep obtained prior to duty start is shorter than six hours, or the duty occurs during the individual’s usual sleep hours (National Research Council, 2011; Williamson et al., 2011). However, in terms of commercial aviation specific safety risks, there does not appear to have been any published systematic attempts to investigate the dose-dependent effects of sleep loss or circadian influences on multi-crew commercial flight performance, and there are limitations to inferring risk standards from aviation accident, incident and safety occurrence databases. It is for this reason that within aviation, sleepiness risk rates associated with duty patterns need to be compared not only against the established laboratory evidence on neurobehavioural performance decline, but also against other parallel safety standards that govern the functional status of the human operator with respect to safe flight performance.

Assessing human factors risks in commercial aviation

Commercial aviation has maintained its record of being one of the safest modes of transportation over the past 40 years, due in large part to substantial increases in the reliability standards of aircraft parts, duplication or triplication of safety-critical sub-systems to prevent single points of failure, strict maintenance schedules, and the ongoing development of high reliability automated control systems (Hartzler, 2014). Commercial aviation is hence considered an ultra-safe high-risk industry (USHRI), as its operations are managed and actively regulated to a high level of safety within an otherwise high-risk environment (Cusick, 2017). In such industries, the accident risk profile usually remains one of extremely low frequency of occurrence, but exceptionally high severity in terms of human and financial costs. Hence the regulatory approach to maintaining high reliability in commercial aviation operations has been to reduce the likelihood of single cause ‘catastrophic’ failures in safety-critical elements of the system to an acceptable minimum.

In commercial aviation, a principle of designing and certificating aircraft for safe flight is that the various sub-systems of the aircraft (the engines, the electrical systems, the pilots, etc.) should meet a quantified reliability standard (Zio et al., 2019). To minimise any risk of a ‘weak link’ in the chain, each sub-system should ideally meet a similar reliability standard. In many cases, in order to achieve this standard, safety critical components of the sub-system are at least duplicated, and sometimes triplicated or more (Tunstall-Pedoe, 1988). Hence, for the sub-system that sustains powered flight, where commercial aircraft have two engines, the design is such that each engine on its own can sustain flight. Similarly, where there are two pilots, in the event of one of the pilots becomes medically incapacitated, the other pilot on their own can continue the flight safely. The inferred impact of pilot failure (partial- or complete) to overall system safety is therefore considered in terms of mechanical system reliability to be at par with the airworthiness requirements of aircraft. However, some hazards are of greater concern because they have the capacity to simultaneously cause duplicated systems to fail, termed common mode failure (Downer, 2009). Examples of common mode failure hazards are flight through flocks of birds or volcanic ash where all the engines may fail at the same time. Thus as part of the overall array of safety-critical sub-systems in aviation, pilots remain central and necessary to the safe operation of modern aircraft (Rosekind et al., 2000) and are also considered to be part of the system that meets a stringent reliability standard. Human physiological capabilities and limitations are accordingly viewed as critical factors affecting the failure rate of the pilot safety sub-system, and so also have associated acceptable safety margins (Roach et al., 2012). A variety of approaches are taken to prevent the failure of the crew component. From a regulatory perspective, the most extensive assurance is provided by preserving the independence of risk between pilots with respect to the rate at which performance between either or both pilots becomes sub-optimal or ceases. An example of this is the practice of ensuring that pilots on the same flight consume different meals prior to and during the flight to mitigate against the possibility of both becoming incapacitated from food poisoning at the same time (ATSB, 2016).

Medical incapacitation rate

The International Civil Aviation Organisation (ICAO) guidance on the medical incapacitation standard for pilots (of no more than 1 occurrence per 1,000,000 hours) is the current acceptable rate of risk for the likelihood of the break-down of optimal or safe performance of the operating flight crew where the cause is 'medically-driven' (ICAO, 2012; Mitchell & Evans, 2004) and represents a probabilistic standard concerning pilot safe functioning in-flight. The rationale of the rule is described as follows (Mitchell & Evans, 2004; see also Figure 2 for a summary);

'The '1% rule' assumes target all-cause, fatal accident rate for large public transport aircraft of 1 per 10^7 flying hours, not more than 10% of which should be due to one system failure (e.g., pilot failure), and not more than 10% of system failures should be due to a subsystem failure (e.g., medical incapacitation). This provides a target fatal accident rate due to aircrew medical incapacitation of 1 accident per 10^9 hours…' (Mitchell & Evans, 2004, p 261).

However, this rate was adjusted upwards (to become more permissible) on the basis of two further assumptions

a) Critical phases of flight constrain the risk to just 10% of total flight time
   The periods where the aircraft is closest to the ground (i.e., take off and initial climb, and approach and landing) are deemed as critical phases, as during these periods '…a slow or incomplete transfer of control is most likely to result in an accident'. At the time the rule was formulated this was assumed as 10% of the total flight time.

b) Existence of second flight crew member
   'In two-pilot operations, which comprise the majority of commercial airline flights, if the handling pilot were to suffer an incapacitation, the other pilot should be able to take control and safely continue the flight'. It was assumed that in only 1 out of 100 events occurring in the critical period would the other pilot not be able to

> take control in time to avoid a fatal accident. This assumption was substantiated by simulator studies which found that in a two-crew operation, in the event of one pilot having a sudden medical incapacitation in a critical phase of flight, in 399 times out of 400 the second pilot could take over the control and continue the flight safely.

Therefore, in order to achieve the target *medical*-cause, fatal accident rate of no more than 1 accident per 10^9 hours (Figure 2, iii), neither pilot should have a risk of medical incapacitation greater than 1 in 10^6 hours, i.e., approximately 1% in 1 year of 8,760 hours (~10,000 hours).

i. Target all-cause fatal accident rate (flight hours)
Frequency of fatal accident – 1e+07
(Target fatal all-cause accident rate)

↓

ii. Fatal accident caused by failure of human component (flight hours)
Frequency of fatal accident – 1e+08
(Single aircraft system contribution < 10% - 10,000,000 x 10 = 100,000,000)

↓

iii. Fatal accident caused by medical incapacitation (flight hours)
Frequency of fatal accident – 1e+09
(Single sub-system (medical incap.) contribution < 10% - 100,000,000 x10 = 1,000,000,000)

↓

iv. Adjustment for critical phase of flight (flight hours)
Frequency of fatal accident -1e+08
(Critical phases of flight ~ 10% flying hours → 1,000,000,000 / 10 = 100,000,000)

↓

v. Adjustment for successful action by other pilot (flight hours)
Frequency of fatal accident – 1e+06
(Co-pilot successful take-over ~ 99% → 100,000,000 /100 = 1,000,000)

Figure 2. Summary of the medical incapacitation standard derivation

This approach was intended to ensure that individuals who are granted a flight crew licence for commercial aviation activities represent a medically fit pilot population, at an acceptably low risk of likely in-flight performance impairment or incapacitation. The medical causes considered as likely to elicit performance decrement that could represent a potential threat to flight safety are a mixed collection of impairments and conditions, which range from sudden serious events such as heart attacks and epileptic fits, through to more mild events, such as headaches and gastrointestinal illness, that still are capable of inducing considerable performance decrements and functional impact to the affected pilot (ATSB, 2007; Evans & Radcliffe, 2012; ICAO, 2012). This collection includes conditions with functional impacts of transient disturbances of attention and/or consciousness such as petit mal epilepsy and common illnesses such as colds, flu and headaches from viral infections. Such incapacitations may hence be due to the effects of a pre-existing medical condition or vulnerability that manifests deleterious symptoms during flight, or the development of an acute medical condition, or physiological event that renders an otherwise fit and well pilot temporarily unable to fly or safely operate the aircraft (ATSB, 2007). Thus medical assessments are predicated on the fact that they address aspects of physical and mental functioning that affect cockpit performance and safety outcomes.

Whilst it is the case that in some regulatory documentation, pilot fatigue, sleeping and ‘microsleeping’ events are referred to as categories of in-flight medical impairments (e.g. ATSB, 2007; Dejohn et al., 2004), in practice sleep and circadian influences on pilot performance are not considered in relation to medical assessment for fitness to fly, except for clinically explicit sleep disorders such as insomnia (Gander & Signal, 2008). Moreover, although there are European and national regulations on pilot duty hours (UK Civil Aviation Authority Authority, 2019), these regulations do not attempt to quantify the acceptable level of fatigue-driven incapacitation risk in the same way as the medical incapacitation risk standard.

Some authors have identified this as ill-advised for a number of reasons. First, it is now well established that sleep and circadian causes of operator fatigue pose serious risks to the safe performance and operation of all modes of transport, and elevated schedule driven fatigue may be predicted from work schedules (Folkard & Åkerstedt, 2004; Marcus & Rosekind, 2017). Indeed, while cardiac incapacitation has not been implicated as a cause of any accidents in a multi-crew cockpit aircraft, ‘pilot fatigue’ has been acknowledged as a causal factor of a number of accidents and serious incidents (Gander & Signal, 2008). Secondly, unlike cardiovascular incapacitation risk, sleep and circadian drive influences on pilots are likely to be similar, since members of crew typically work the same duty pattern. This means there is a diminished likelihood of the independence in fatigue-related risk causing human break down in performance. A particular example of this concerns involuntary sleep as a form of double mode failure, as electroencephalography (EEG) and survey evidence suggests that simultaneous severe sleepiness and sleep events do occur in both flying pilots during normal operations (Cabon et al., 2003; Cabon et al., 1993; ComRes, 2013; Samel et al., 1997), may occur without the pilots' awareness (CAA, 2003a; Graeber et al., 1990) and incident report data suggest that pilot fatigue has affected both operating flight crew on a number of occasions, highlighting the vulnerability of flight crew to simultaneous fatigue-related incapacitation risks (e.g. NTSB Reports: 2017: NTSB/AIR-18/01, pp52-53; 2009: NTSB/AAR-10/01, pp106-107). More broadly however, as described earlier in Section 1.4, sleep-related incapacitations affect sensory, cognitive, physical and behavioural functioning of crew and thus are likely to be functionally similar in terms of severity of impact to a variety of medically-driven incapacitations. Finally, as other authors have discussed, operator fatigue constitutes a more insidious and common source of pilot impairment (Caldwell, 2005; Gander & Signal, 2008), which might be more dangerous than an obvious, complete incapacitation if the impairment in performance goes undetected and the human component operates at an unsafe level for an extended period of time. Thus, in terms of consistency within safety systems, it follows that acceptable risk rates of flight crew performance decrements and incapacitations due to sleepiness should be considered within the same risk framework as medical causes, even though this is not current practice. Safety concerns around pilot fatigue furthermore appear to be shared by UK aviation medical examiners, who are in charge of the medical certification of pilots' licences. In 2011, interviews conducted with 42 UK registered Aviation Medical Examiners (AMEs) revealed that 66% considered fatigue as one of the two systems areas with the greatest potential to cause an airline accident (ComRes, 2016). A majority of AME's (82%) further felt that particular safety problem of pilot fatigue is that it can affect both pilots of the aircraft at the same time. The medical incapacitation rate hence provides a useful aviation specific benchmark for appraising how sleepiness occurrences and predicted fatigue risk rates associated with pilot rosters may compare against other risks to operator safe performance.

Alcohol limitation standards

A further standard relating to the acceptability of risk to safe pilot performance is that concerning alcohol-related impairment. Under both EU regulations (EU 965/2012 CAT.GEN.MPA.100 (c)) and UK regulations subsequently retained and applied from 14th August 2018 (UK Reg EU 956/2012; EU 2018/1042), flight crew members are not permitted to perform duties on an aircraft when under the influence of alcohol, which has a threshold level of 20 milligrammes of alcohol per 100 millilitres of blood, otherwise referred to as a blood alcohol concentration (BAC) of 0.02% (Civil Aviation Authority (CAA), 2022). This threshold is a quarter of the road user alcohol limit of 0.08% used in national driving laws. BAC is hence a metric used commonly in legal transport settings and can be useful to refer to in terms of public policy as most adults have an understanding of the effects of alcohol intoxication on cognitive performance.

The aviation BAC standard is of theoretical and practical interest because it is lower than that permissible in driving, and previous research has suggested that sleep loss and alcohol intoxication show similar effects on aspects of cognitive performance (e.g. psychomotor performance) that are relevant to transport safety settings (Dawson & Reid, 1997; Lamond & Dawson, 2002; Williamson & Feyer, 2000). Here 'psychomotor' performance refers to attention, vigilance, and precise motor responses on particular tasks. Some studies have suggested that as continuous daytime waking exceeds 16 hours, performance deficits begin to increase to levels equivalent to blood alcohol concentrations (BACs) between 0.05 and 0.1% (Williamson & Feyer, 2000; Dawson & Reid, 1997; Lamond & Dawson, 1999). In Dawson and Reid's (1997) study involving 40 participants taking part in two counterbalanced experiments, performance on a hand-eye tracing and coordination task was reported to deteriorate in a linear manner for both the condition involving 28 hours of sleep deprivation, and the condition with increasing BAC. By 17 hours of continued wakefulness, subjects' decline in performance was reported to show equivalence to a BAC of 0.05%, and by 24 hours this increased to a BAC of 0.10%. The interpretation of these alcohol equivalence findings studies have been critiqued on analytical grounds mainly in terms of the findings around the magnitude of impairment (Maruff et al., 2005). For example, in Dawson's studies, for each level of fatigue and % BAC, the magnitude of impairment was expressed using only a percentage change score in which the group mean performance at each trial was subtracted from the group mean score at the related baseline trial and expressed as a percentage. This means that the normal variability in performance across different trials was not taken into account (Maruff et al., 2005). A separate criticism was that these studies did not sufficiently appreciate the larger within-group variability in cognitive performance with both increased % BAC and sleep deprivation. However, in addressing these limitations Maruff's study applied a measure of effect size to express the magnitude of performance impairment associated with increasing fatigue and % BAC that took into account differences between group means as within-group variability. In making this correction the magnitude of the impairment and variability of psychomotor performance detected at 24 hours TSD was slightly greater than the impairment observed in the same individuals at 0.05% BAC, but there was no level of sustained wakefulness that equated the psychomotor performance detriment associated with 0.08% BAC. Thus the magnitude of sleep-deprived impairment was not as great as estimated in previous studies. However, in line with previous studies, increasing levels of fatigue and alcohol appeared to have qualitatively and quantitatively similar effects, particularly in terms of increasing reaction times and decreasing the consistency of psychomotor performance. The equivalence of this impairment could mainly be seen through performance variability reaching equivalence to 0.02% BAC by 14 hours and 0.05% BAC by 24 hours (Maruff et al., 2005).

The relevance of the 0.02% BAC standard in aviation is that extended work hours (due to operational needs, delays etc) have been commonly associated with pilot schedules, which consequently has an impact on the number of hours pilots spend continually awake that may extend up to 20 hours (Caldwell & Caldwell, 2003; Rosekind et al., 1994; Gander et al., 1998). Together these studies hence provide a basis for comparison and suggest that fatigue-related impairment in psychomotor performance may be comparable in some important ways to alcohol-related impairment that has relevance for the consistency of safety protection for human performance.

Fatigue regulations

The legal landscape covering operator and crew responsibilities regarding fatigue, pilot hours and safety reporting is made up of laws and guidance that specifically relate to pilot fatigue, and those that relate to aviation safety and occupational safety more generally. Although the UK is now in a period of transition from EU regulations, pilot and operator responsibilities in current practice still reflect the obligations as specified by the European Safety Agency (EASA) under Commission regulation (CAA, 2021b; Commission Regulation (EU) No 83/2014, 2014). For regulations specific to pilot fatigue, there is essentially a tripartite structure. There are overarching principles that state that operators should not construct fatiguing rosters (ORO.FTL.110 Operator responsibilities) and that flight crew members should not fly if they suspect they may be suffering from fatigue that could endanger the safety of the aircraft (ORO.FRL.115 Crew member responsibilities). There are a set of prescriptive regulations (ORO.FTL.205- 245) that delimit the maximum flying and duty hours pilots may operate for particular time periods, depending on the time of day, whether the crew is ‘acclimatized’ (their body clock matches the local time zone) and number of ‘sectors’ (flights). Finally there are regulations around fatigue risk management (ORO.FTL.120) which outline the expected principles of how organisations should establish, implement and maintain fatigue risk management processes based on scientific principles, hazard identification and risk assessment and act on operational risk data.

ORO.FTL.120

> *(a) ‘….the operator shall establish, implement and maintain a FRM as an integral part of its management system...*
>
> *(b) ‘ The FRM established, implemented and maintained shall provide for continuous improvement to the overall performance of the FRM and shall include: …’*
>
> > *(3) ‘scientific principles and knowledge’*
> >
> > *(4) ‘ a hazard identification and risk assessment process that allows managing the operational risk(s) of the operator arising from crew member fatigue on a continuous basis;’*
> >
> > *(5) ‘….a risk mitigation process that provides for remedial actions to be implemented promptly, which are necessary to effectively mitigate the operator’s risk(s) arising from crew member fatigue and for continuous monitoring and regular assessment of the mitigation of fatigue risks achieved by such actions.’*

In the terminology of FTL regulations, particularly for prescriptive regulations (ORO.FTL.205-245), key terms have neither their plain English nor their scientific meaning. For example, ‘maximum flight duty period’ is not a length of time that can never be exceeded, as to this ‘maximum’ are various extensions and discretionary periods that can be added on the duty day. In this regard, the phrase ‘basic un-extended maximum’ is sometimes used. The unconventional use of terms and the complex architecture of the rules, where several rules have to be cross referenced in order to capture all those relevant to a given situation, make FTL regulations difficult to comprehend to any individual not well-versed in this area. However, implied by the tripartite structure is the recognition that prescriptive regulations on hours-of-work limits alone cannot control for the potential for fatiguing rosters to be constructed. Indeed, as demonstrated by the 2017 Air Canada incident (that had pilots flying at over 19 hours awake after being called from standby), roster patterns that comply with national prescriptive regulations do not always protect pilots from schedule-driven fatigue risks (NTSB, 2017). Further highlighting this vulnerability is that over 40% of fatigue-related safety recommendations in aviation produced by NTSB investigations have related to aviation scheduling policies (Marcus & Rosekind, 2017). The prescriptive approach further has shortcomings in terms of the controls available to deal with the dynamic hazard of fatigue. For example, whilst there are minimum rest periods seeking to control acute sleep deprivation levels of pilots, rest periods of identical numerical length will not have the same recovery value across the twenty-four hour cycle or between different work schedules, since the propensity for sleep is heavily influenced by circadian rhythms. As such, there are limits to the efficacy of traditional hours-of-service regulations for controlling the hazard of fatigue where many variable shift-work patterns and forms of circadian disruption are possible, and the interactions of such factors are not easily summarised into specific prescriptive rules (Cabon et al., 2008). Thus, in order to address some of the limitations of such prescriptive regulations, in recent years there has been an increasing regulatory drive towards encouraging greater proactive safety management of fatigue risks, in accordance with the principles outlined in ORO.FTL.120. Formal systems developed for this purpose, referred to as Fatigue Risk Management Systems (FRMS), represent a subset of proactive safety management policies and procedures covered under an organisation's Safety Management System (SMS).

Safety and Fatigue Risk Management Systems (SMS & FRMS)

Safety Management Systems represent process-based approaches to managing safety which are undertaken by the organisations that generate the risk. Collectively SMS seek to establish a set of proactive, systematic activities around safety policy formulation, hazard identification, tailored risk assessments, collection and interpretation of operational data and the development of risk mitigations, intended as a cycle of continuous safety improvement processes. At a general level, such systems appear most frequently concerned with industrial safety, in for example, transport and nuclear settings, where risks arise within the context of occupational activities. Most definitions of SMS implicate the need for structured processes and management responsibilities at different organisation levels to improve the safe conduct of work. Within the aviation context, ICAO (2012, p12) defines 'safety' as '*the state in which the possibility of harm to persons or of property damage is reduced to, and maintained at or below, an acceptable level through a continuing process of hazard identification and safety risk management.*' Against this concept of safety, SMS is defined as '*an organized approach to managing safety, including the necessary organisational structures, accountabilities, policies and procedures'* (ICAO, 2012, p11). Fatigue Risk Management Systems (FRMS) represent a specific development of SMS for the systematic identification of workplace hazards relating to fatigue (Sprajcer, 2022). As such, there is a strong rationale supporting the principles of fatigue risk management systems, in so far as they are intended to be data-driven, permitting ongoing evaluation of operational experience against established risk thresholds. However, within commercial aviation, FRMSs are also defined as a '*scientifically-based, data-driven flexible alternative to prescriptive flight and duty time limitations… [involving] a continuous process of monitoring and managing fatigue risk'* (Graeber, 2008). Thus FRMSs are considered not only as a potential complement to prescriptive rules controlling the hazard of fatigue, but also more ambitiously as an alternative means of compliance for companies to expand their operational productivity through demonstrating the maintenance of or 'equivalent level' of safety on the basis of operational data (Cabon, 2008). There are arguably several components of FRMS at various layers of abstraction that may feed into a company's SMS at an organisational level. Those identified as the central six components by representatives from IATA, ICAO and IFALPA (2011, p13) include;

1) *Policies* (scope definition, management commitment, shared and individual responsibilities, safety objectives and activities)
2) *Documentation* (processes and procedures, mechanisms for stakeholder involvement, records, report outputs and actions)
3) *Risk management processes* (fatigue hazard identification, risk assessment and risk mitigation)
4) *Safety assurance processes* (performance monitoring, management of operational changes, continuous improvement)
5) *Safety promotion processes* (training programs and communication plans)
6) *Fatigue Safety Action Group* (coordination of fatigue risk management activities)

Against this context, the underlying fundamental notions implied by definitions of safety and SMS as set out above, alongside basic components of FRMSs are that 1) risks cannot be reduced to zero but rather to agreed acceptable levels and 2) the safety of given activities must be measured and evaluated against such established acceptable levels. As Stolzer (2016, p 20) notes, the SMS approach means that '*…for safety professionals the very word "safety" implies constant measurement, evaluation and feedback into the system'*. As set out at the beginning of 1.8, however, a central scientific issue relating to the SMS approach when applied to the hazard of fatigue in the form of FRMS is that there is no defined acceptable level of ‘fatigue’ or ‘sleepiness’ for safe flight, the colloquially termed ‘*how tired is too tired?*’ problem for practitioners. This is for two reasons: first, the complex challenge of marrying the advances from sleep science on how fatigue impacts neurobehavioural performance variability with more limited and complex operational evidence implicating ‘pilot fatigue’ as a risk factor to whole-system safety in commercial aviation; second, the lack of data on the dose-dependent effects of sleep loss or circadian influences on multi-crew commercial flight performance. However across industry, FRMS policy, risk management and safety assurance processes require the establishment and communication of acceptable levels of safety with respect to the hazard of fatigue. It is for these reasons that the case has been made in Section 1.8 that within commercial aviation, sleepiness risk rates associated with duty patterns need to be compared not only against the established laboratory evidence on neurobehavioural performance decline, but also against other parallel safety standards that already exist within commercial aviation to govern the functional status of the human operator (or safety-critical sub-component of the system) with respect to safe flight performance. These may be particularly important to consider where organisations are encouraged to self-regulate on risks where biasing factors may be elevated (such as fatigue controls directly limiting commercial productivity), and there are organisational incentives relating to financial performance (Stolzer, 2016; p25). Indeed, various authors have issued concerns around FRMS implementation in practice (Gander et al., 2011; Gander, 2015; Sprajcer et al., 2022). First, the implementation of such systems shifts the locus of responsibility for safety away from the regulator towards companies and individuals, requiring far greater scientific understanding from traditional safety management roles and resources than may be readily available (Gander et al., 2011). This means that there is an organisational vulnerability around the selection of appropriate data measures and acceptable thresholds for FRMS that may not be sufficiently informed by the latest developments in sleep science. Secondly, effective implementation of such systems is likely to depend substantially on the existing organisational safety culture, including fatigue reporting from staff which may or may not be reliable (Sprajcer et al., 2022). Indeed, industry reports suggest that in contrast to generally positive perceptions of flight crew towards safety culture within commercial aviation, the same sentiments are not held for the hazard of fatigue (Reader et al., 2016). In general, fatigue, although widely experienced by flight crew whilst at work, is underreported via formal safety channels, and crew may often continue to work even with severe symptoms of fatigue (*see sections 1.6 and 1.9)*. Since feedback from flight crew into FRMS (as the crucial input to data-based organisational learning) is mostly based on subjective fatigue reports, reporting culture is crucial to the success of the FRMS. As such, whether or not an FRMS is primarily viewed and treated as a productivity enhancement tool or a protective system is likely to impact the trust levels of staff working within an airline and underlie its success of the FRMSs generating an ‘equivalent or greater level of safety’ of the broader SMS (Steiner et al., 2012). However, data collected from flight crew across Europe within the last decade suggests that pilot perceptions of safety culture around fatigue management are not positive (ECA, 2012; Reader et al., 2016). The *Future Sky Safety* report, a collaborative research exercise conducted by representatives from London School of Economics (LSE), Eurocontrol and Netherlands Aerospace Centre (NLR), collated survey responses from 7,239 (14%) of the European flight crew population found that whilst perceptions of safety culture are generally positive, for example almost all pilots (93.49%) agreed that their colleagues are committed to safety, over half of the sample (50.05%) felt that fatigue was not taken seriously within their organisation. The authors moreover concluded that fatigue survey items within the broader safety culture questionnaire were responded to in a consistently negative fashion (Reader et al., 2016, p17). Thus there are legitimate concerns both from first principles and survey findings that without a widely accepted measuring method and associated thresholds, nor evidence of effective safety reporting culture around fatigue, there are no guarantees that FRMSs lead necessarily to equivalent or greater levels of safety (Steiner et al., 2012). Thirdly, self-determination of risk at the discretion of the operator is not without its own vulnerabilities, as conflicts between productivity and safety are inherent organisational risks that may be more difficult for operators to objectively appraise (Gander et al., 2015; Hunter, 2015). Indeed, there is substantial historical precedent for caution around the safety assurance provided by safety management systems. The phrase ‘*You’ve got to draw the line somewhere’* was coined during parliamentary debates prior to the 1876 Merchant Shipping Act, where ship owners were originally judged to be best placed to determine how heavily loaded their ships would be. Following continuing loss of lives at sea from overloaded ships driven by economic pressures, the original ‘SMS’ around load lines (and therefore judgement of risk threshold for the loading capacity of ships) was eventually decided to be better determined by an independent body, and thus the Plimsoll line was established (Hunter, 2015). Such examples from history amongst others in more recent years – Deepwater Horizon oil spill and Three Mile Island disaster - show how forms of self-regulation can easily evolve from alternative means of protection into production enhancing tools based on organisational financial incentives that increase risk over time (Steiner, et al., 2012). Hence for modern commercial aviation, continued development of consensus on acceptable fatigue risk thresholds and the consistent measurement of fatigue risks represents an essential ongoing set of independent research activities. This is not only to inform good practice for individual company FRMS activities, but also more importantly to ensure that at an industry level, the collection of fatigue risk data is not confined just to serious incidents but also includes coverage of normal operations so that findings may be appraised against common set of fatigue risk measures and risk denominators. In the absence of an evidenced-based consensus on such metrics, the intended value of the entire FRMS and SMS framework for commercial aviation is arguably severely weakened.

1.9 Incidence of Fatigue and Sleep-related phenomena in commercial aviation

At a coarse level, there have been a few attempts within some surveys to identify the rates with which high levels of sleepiness or ‘pilot fatigue’ are experienced. Survey studies suggest that in-flight fatigue is a widespread issue within commercial aviation, with prevalence data suggesting in-flight fatigue has been experienced by the majority (69- 91%) of commercial airline pilots at some point within their careers (Aljurf et al., 2018; Jackson & Earl, 2006; Reis et al., 2016). In one study, data collected from a sample of long-haul pilots indicated that nearly 66% of pilots experienced significant fatigue at least once a week, and more than 96% reported that fatigue had interfered with their usual social activities during the last month (Petrie et al., 2004). Another showed that 43% of UK airline pilots (n=215) from both long and short haul carriers believed their abilities had been compromised at least once a month by tiredness, with 84% reporting this had occurred during the past 6 months (ComRes, 2013). At the European level, survey polls with commercial pilots reveal similar high percentages. In Germany, 92% of pilots felt ‘too tired’ while on the flight deck at least once in the past three years, and for the same question, similar high percentages have been reported for Austrian airline pilots (85%), Swedish pilots (89%), Danish pilots (93%) and 67% of pilots from the Netherlands (European Cockpit Association, 2012). Instances of involuntary sleep on the flight deck are also widely reported among commercial airline pilots, with estimates ranging from 56% to 71% of commercial airline pilots having experienced it at some point in their careers (ComRes, 2013; Rosekind et al., 2000). In terms of simultaneous sleep-driven incapacitation, almost a third of UK airline pilots who had reported having involuntarily fallen asleep on the flight deck also reported to have also had an occasion where they woke up to find the other flying crew member had fallen asleep (ComRes, 2013). Taken together, these survey data suggest that pilots suffer from high levels of sleepiness during flight, and that a large proportion of crew may have experienced involuntary sleep events.

Whilst such studies offer some overarching insights on the burden of pilot fatigue across the wider national and international commercial aviation industry, it is difficult to determine comparable incidence rates from the findings of such studies due to the different conceptualisations of fatigue used, and absence of similar time scales involved with respect to operational variables. As such, there is a considerable lack of research data on more detailed estimations of how frequently high levels of sleepiness or sleep events occur during normal commercial aviation operations. Previous studies furthermore do not directly assess the frequency of involuntary sleep events during normal operations, when these occur, nor establish a rate of any fatigue-related events per flight hour which are important metrics for understanding risk and threats to safety.

One study that did attempt to offer some evidence in this regard was conducted by Houston, Dawson and Butler (2012). These authors reviewed the voluntary fatigue reports submitted by pilots and cabin crew within a short- and medium-haul airline over a period of 12 months. The authors determined the annual crude incidence rate of fatigue reporting to the company of 103 cases per 1000 persons per year (for pilots). To the best of knowledge, this appeared to be the only published peer-reviewed study that attempted to assess a form of incidence rate beyond survey data. Presumably due to confidentiality reasons, a number of key factors such as the location, size and operational shift types, and number of flight hours conducted per year by this airline were not reported, limiting the interpretation of their findings and their generalizability. Setting these limitations aside, it should be noted that Houston's study also focussed on voluntary reports submitted by pilots to the company (Houston et al., 2012). Whilst these certainly constitute important operational data, using such reports to approximate incidence rate and severity of on-duty sleepiness is potentially problematic. As mentioned previously, confidential research with pilots has suggested that as a group there is a large degree of underreporting of fatigue from pilots via formal channels to their own company (Confidential Incident Reporting Programme (CHIRP), 2017; Caldwell, 2001; Reis et al., 2013). Indeed, despite survey data indicating high levels of involuntary sleep on the flight deck within British Airline Pilots, in terms of regulatory data, just two reports of involuntary sleep were submitted to the UK aviation regulator between 1976 – 2013 (BBC Freedom of Information Act request to the CAA: F0001485, 2013). Hence, formal safety reports to companies likely underrepresent the real world incidence levels. Other research involving the monitoring of pilot fatigue or sleepiness levels during actual operations has also typically focussed on addressing a very specific set of operational issues for a particular airline or type of operation (Gander et al., 2013; Samel et al., 2004; Signal et al., 2013; Wright & McGown, 2001). As a result, the number and range of participating pilots or types of flight assessed from different airlines have understandably, tended to be somewhat restricted. Previous field and survey studies have thus not produced an overarching picture of both reported and predicted incidence rates of sleep related phenomena or high sleepiness levels occurring during flight, which are important metrics for understanding the risk exposure and assessment of the threat to safety across the aviation industry.

1.10 Sleep Science and recent developments in Safety Theory

As detailed in earlier sections of this chapter, major advances in the applied field of sleep science have involved developments in the prediction of wake-time sleepiness levels (through bio-mathematical modelling of work hours), coupled with the measurement of the spectrum of alertness and sleepiness phenomena that may be experienced by shift workers whilst on duty (through both validated subjective scales and simple tests on cognitive performance). Such data insights have helped to inform safety-related risk assessments within commercial aviation by highlighting the dynamic nature of fatigue risks and the limitations of traditional working time limits for controlling the hazard of fatigue in organisational settings that operate complex shift-work schedules. The acceleration of such approaches can be traced back to a period of extensive research efforts between 1980s-2000s that saw large scale data collection efforts and validation work on Borbély’s 2-process model (Borbély, 1982; Borbély & Acherman, 1999; Dinges, 2004; Mallis et al., 2004). This was coupled with the increasing recognition at the regulatory level during the early 2000s that modelling and measurement approaches represented important upstream components of the FRMS toolbox available for organisations to adopt to proactively assess risks during normal operations, alongside the traditional forms of safety learning from adverse safety occurrence reporting through post-flight fatigue reports (CASA, 2010; 2014). Thus the combination of proactive and retrospective approaches to evidence gathering has become established within commercial aviation over the last two decades. This includes the emphasis on monitoring not just fatigue-related safety occurrences, but also fluctuations along the alertness-sleepiness continuum for the occupational hazard of fatigue. As such, advances in applied sleep science research have complemented conceptual developments within broader safety theory and management in recent years, and indeed in a number of areas may have anticipated such developments.

Over the last two decades, some safety management theorists have posited the need for a conceptual shift from traditional definitions of ‘safety’ that appear to measure its presence indirectly in terms of the absence of negatives or adverse outcomes (near misses, incidents, accidents) towards definitions that reflect how everyday work elicits acceptable or even successful outcomes. As conceptualised by Hollnagel (2015; 2018) a coarse description of this theoretical distinction is that the ‘*Safety I’* traditional approach to measuring safety focuses on event outcomes of things that ‘go wrong’ with a retrospective focus on the analysis of failure, whilst ‘*Safety II’* approaches attempt to assess and draw attention to things that ‘go right’ and to ‘learn from what works’ (Hollnagel, 2015, p 28). As such, with respect to organisational practice it is argued that *Safety I* principles promote a reactive organisational response to events and a focus on lagging indicators of safety, whereas *Safety II* principles imply continuous monitoring of everyday operations and proactive surveillance of how performance variability under different conditions generates ‘high frequency, acceptable outcomes’ (Hollnagel, 2015, p 25). The particular relevance of this distinction in safety theory to industries such as commercial aviation, as argued by Hollnagel and others (e.g. Provan et al., 2020), is that *Safety II* principles facilitate organisational learning about safety through their inclusive focus on the entirety of operational performance data rather than a concentrated focus on single event failures. Since the frequency profile of accidents amongst ultra-safe high-risk industries has substantially reduced due to technological advances over the past few decades, a focus on single adverse events correspondingly no longer yields sufficient data for systematic analysis of potential vulnerabilities in operational performance and associated control mechanisms. Thus, without the adoption of a *Safety II* approach, it is argued, the data available for safety learning is severely diminished (Provan et al., 2020). At the same time, such theorists argue that performance demands within modern organisational systems are significantly greater, with systems more complex and highly interdependent, and that this further limits the value of safety learning that can be obtained from ‘monocausal’ or ‘composite’ linear explanations (Dekker et al., 2011; Hollnagel, 2015). In parallel, it is argued that the traditional *Safety I* approach encourages inappropriate focus on human error as the default explanation rather than assessing weaknesses across the entire socio-technical system and their interaction across multiple organisational layers. A third major difference of approach, according to such theorists, lies in the distinction between the

Safety I philosophy of regarding the difference between failure and success as bimodal – either a system works correctly or it does not – and the *Safety II* approach that asserts a commonality, or equivalence, between the factors underlying success and failure (Hollnagel, 2015; 2018). This *Safety II* principle rests on the fundamental assumption of human performance adjustability; i.e. that systems work for most of the time because people can detect emerging problems and take corrective action before a situation deteriorates. Performance variability is thus argued to be inherent in the functioning of any socio-technical system, beyond the simplest. Since it theoretically cannot be eliminated, it has to be proactively managed and dampened where it might lead to adverse outcomes. However, within the framework of *Safety II*, the underlying assumption is that performance variability within a system is seen as essentially beneficial, particularly where humans intentionally adapt to different circumstances that may not be anticipated by work 'as planned' to achieve safe outcomes (Provan et al., 2020). Findings from sleep science, however, highlight at least two other possibilities: first, pilots may achieve a successful flight outcome despite one or more of the flight crew knowingly experiencing high fatigue because they intentionally make exceptional efforts to remain alert and combat the higher probabilistic risk of pilot errors; alternatively, pilots may achieve a successful flight outcome despite being unaware that they are suffering extreme fatigue, and thus have been reliant on the aircraft's extensive supplementary safety systems and other safety netting processes to achieve a safe landing. In each of these cases of performance variability, the safety margin relating to having one or two optimally functioning flying pilots in the cockpit has been heavily compromised, and not in ways that can be logically argued to benefit safety. What these alternative possibilities demonstrate is that measurement of performance variability in relation to the safety hazard of fatigue needs to be carefully distinguished from overall measurement of safe flight outcomes. Furthermore, the idea of beneficial performance variability critically also presumes intact functioning for adaptive context-dependent responses to produce safe outcomes. Sleep scientific insights would heavily imply that any manifestations of fatigue-driven performance variability (for example, around standard operating procedures during critical phases of flight) by contrast are likely to degrade essential system safe guards rather than provide adaptive utility.

Notwithstanding these caveats, a focus on performance variability is a point of evident convergence of approaches between these conceptual propositions within safety theorising and the way that methods from sleep science have informed the design and practical application of Fatigue Risk Management Systems for controlling the hazard of human fatigue. These are specifically designed not to eliminate the risk of pilot fatigue, which is an inevitable reality of commercial aviation operations due to the impact of shift work on sleep and circadian drives. Instead the aim is to reduce as far as possible the incidence of elevated fatigue levels that may potentially contribute to adverse operational outcomes, and to not proceed with operations with fatigue risks if they exceed certain threshold levels. As set out in Table 2, specifically in relation to pilot fatigue, many of the advantages proposed by Safety II core principles (Hollnagel, 2018) are already well-established components of applied sleep science methodology and existing FRMS design principles already encompass a hybrid blend of both Safety 1 and Safety 2 approaches (IATA, 2014).

Table 2. Comparison of Safety I and II theoretical propositions against sleep science methods applicable to commercial aviation practice

Safety I	Safety II	Applied sleep science methods in commercial aviation and FRMS principles
Single event focus, based on severity of occurrence (Reactive activity, focus on accidents and incidents, search for root causes valuable but wider learning opportunities limited)	Wider timespan and scope of analysis on everyday actions and outcomes (Proactive activity, focus on normal operations or more probable events, greater overall learning opportunities)	Focus on risk profiles within normal operations *and* learning from safety occurrences. Data inputs can be proactive and reactive activities including: *Pre-flight risk* - e.g. Probabilistic risk estimates from modelling of rosters - e.g. Analysis of sleep diaries - e.g. ‘on-the-day’ PVT tests for vigilant attention or subjective sleepiness ratings *On-duty risk* - e.g. top of descent self-assessment of sleepiness ratings - e.g. feedback from staff (including refusal to operate next flight / extend duty due to fatigue) *Post-flight learning* - e.g. Capture of subjective fatigue and safety experience through fatigue reports - e.g. Reviews of safety occurrence data where fatigue highlighted as symptom or contributing factor
Focus on human error, not on weakness of socio-technical system	Extension of analysis across different organisational layers	FRMS is theoretically designed to facilitate continuous iterative feedback and analysis from different organisational layers - e.g. feedback from staff involved in rostering, planning and monitoring safety data as well as pilots at FRMS meetings, which in turn is fed into a company’s SMS on wider system safety - e.g. concurrent review of objective predictive modelling of pilot rosters and subjective

		experience captured in fatigue reports - e.g. review of adverse occurrence data where fatigue cited or otherwise identified as a contributory cause, assessment against other safety occurrence data
Bi-modal classification of systems as either either safe or not	Learning from degrees of performance variability and a spectrum of safety outcomes	Data informing safety impact of fatigue risk can include; - Average fatigue risks associated with normal operations (affecting entire work force). - Routes and rosters with increased fatigue risk (predicted and reported) - On-duty sleepiness symptoms experienced by staff denoting increased probability of performance variability (increasing unsafe performance at the individual pilot level) - On-duty sleepiness symptoms experienced by staff alongside detriments in performance with operational consequences (human sub-component and wider impact on other aircraft systems)

Whilst, as noted in Section 1.8, there may be limited evidence of how well FRMS are implemented in practice, it is important to note that FRMSs do not reflect a simplistic bimodal classification of fatigue risk. Instead, the roster design adopted by many aviation operators is informed by approaches (such as bio-mathematical modelling) that generate probabilistic assessments of fatigue risk along a spectrum, with scope for pilots, in certain circumstances, to agree to or override extensions to flight hours through use of so-called 'Commander's discretion' (EASA: ORO.FTL.205(f)). Although, once again, the success of such operational practices for maintaining safety may critically depend on the existing safety culture within the airline, this is an illustration of how performance adjustability around the safety hazard of fatigue can in theory operate. However, the unresolved safety issue with FRMS design is that there is no consensus among commercial airline operators on what constitutes an acceptable normal range of fatigue, nor on how sleepiness levels should be measured. The principles of *Safety II* do not overcome this problem, nor do they provide a practical way to resolve it, because agreed measurement techniques, common risk denominators and methods of comparison with other SMS risks are either presupposed or not specified. It is precisely these risk appraisal and measurement challenges that are the central focus of this thesis. Indeed, a primary issue that emerges from this discussion of *Safety I* and *Safety II* models alongside findings from sleep science is the need to distinguish between operational success, defined as a safe flight outcome, and, as a prior condition, the safe management of fatigue, defined as the maintenance of fatigue levels within an acceptable maximum limit. Focusing safety measurement on flight outcomes without measuring the preceding fatigue levels experienced by pilots risks masking a serious erosion of the invisible safety margin on fatigue. A flight that lands without operational incident but with pilots experiencing levels of fatigue elevated well above maximum acceptable limits represents a weak link in risk controls related to the human factor.

Conclusion

Operator fatigue remains a prominent safety concern in commercial aviation as schedule features inherent in pilot work patterns lead to circadian desynchrony, sleeping difficulties and on-duty sleepiness. A considerable body of sleep science research has demonstrated that both total and partial forms of sleep deprivation cause sleepiness, a brain state associated with a wide array of cognitive and performance decrements and involuntary sleep events. Pilots remain central and necessary to the safe operation of modern commercial aircraft and are considered in engineering terms as safety-critical components, with associated safety margins for human physiological capabilities and limitations. However, in contrast to most medical forms of partial or total incapacitation, pilot fatigue risks may not be assumed to be independent between flight crew undertaking the same work patterns. Safety investigators and researchers also have identified fatigue as a factor affecting both operating flight crew in serious incident reports, highlighting the vulnerability of the human component to simultaneous fatigue-related incapacitation risks and a potential greater impact to overall system safety.

The practitioner-based question of ‘how tired is too tired for safe flight?’ represents an important research need for industry, given the increasing regulatory emphasis on performance-based regulation of fatigue risks. This question has been historically difficult to address for a number of reasons; 1) there have been no systematic sleep deprivation studies addressing multi-crew flight performance, 2) prior research on fatigue risk exposure levels drawn from safety report databases may have considerable limitations and 3) there has been limited research assessing incidence and severity of in-flight sleepiness during normal operations. This thesis seeks to provide new evidence to address some of these important evidence gaps in the existing research literature and advance a theory-based approach for improving data collection and the appraisal of fatigue risk exposure within commercial aviation.

Chapter 2. Methods

2.1 The theoretical approach

The Fatigue risk trajectory or 'defences in depth' model (FRT, Dawson & McCulloch, 2005) is a conceptual framework that provides an approach for appraising fatigue risks that may contribute to the breakdown of safe operator performance and broader system safety. This hierarchical framework is discussed below for its relative merits as a superior diagnostic model for structuring fatigue risk management systems and incorporating insights that align with both *Safety I* and *II* thinking. While it owes much of its conceptual foundations to Reason's seminal 'Swiss Cheese' model (SCM) of accident trajectory (Reason, 1990, 1995, 2000), it builds on Reason's crucial highlighting of distal and proximal defences and has subsequently developed from this into a framework that encompasses proactive and reactive assessment of accumulating fatigue risks across specific organisational layers and operational points in time, and provides scope for greater precision on leading and lagging indicators for fatigue-related safety occurrences.

Under Reason's early iterations of the SCM model (Figure 3a), the safety defences in an organisational system are depicted as consecutive slices of cheese, each providing independent layers of protection against a particular safety hazard. The swiss cheese 'holes' of varying size and position in each layer represent the weaknesses of any given organisational defence. The theoretical proposition illustrated by Reason's SCM is that accidents occur as a result of holes (weaknesses) within each defence layer aligning, permitting a trajectory of accident opportunity (depicted as the red arrow). Organisational systems which are considered 'ultra-safe' hence depend on having multiple defence layers, which collectively reduce the likelihood of a single point of failure (where all the holes align) resulting in a whole-systems breakdown from a particular hazard.

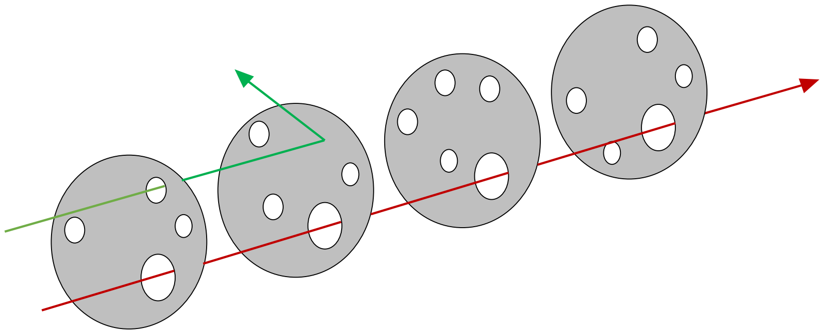


Figure 3a. A basic illustration of Reason's (1990, 2000) Swiss Cheese Model of accident opportunity. The circular slices represent protection or defence barriers, and the holes as failures or active errors within these. The arrows represent paths leading to the risk of an accident

Expanding on this basic principle, Reason's further depictions of accident causation (Figure 3b) highlight the theory that it is the combination of various failures in distal organisational defences as well as proximal unsafe acts by frontline operators that together pave the way for a whole-systems accident opportunity.

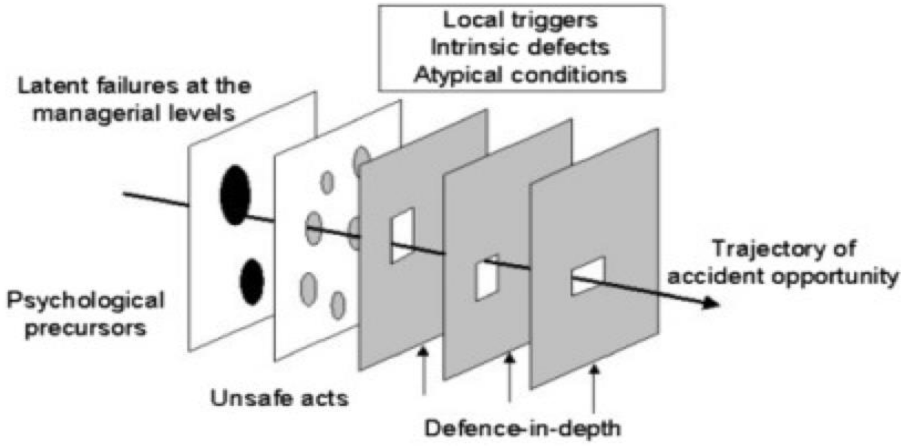


Figure 3b. The Swiss Cheese Model highlighting Reason's (1990; 2000) systems focus on accident causation, illustrating that distal organisational failures may combine with proximal conditions and unsafe acts of frontline workers to breach safety defences.

A clear implication of the SCM model for applied field safety research is the value of identifying and quantifying the weaknesses in these organisational defence layers for particular hazards, such that the rate at which they may fail and the conditions under which they may fail can be better understood, and either mitigated or reduced, or new defences added to better combat the hazard.

Within the field of transport safety, Reason's conceptual Swiss Cheese model has for many years provided a powerful metaphor of accident causation. In particular it emphasizes the need to consider both distal and proximal risk factors that may interact to endanger whole-system safety from a particular safety hazard (Larouzee & le Coze, 2020). Whilst many practitioners have intuitively embraced the illustration aids and interpretive flexibility of the Swiss Cheese Model, it has attracted a number of criticisms (e.g. Dekker, 2002; Hollnagel, 2004; Luxhøj & Maurino, 2001), including from Reason himself (Reason et al., 2006) who identified its irrefutability by standard scientific methods**.** Some authors have viewed the SCM's simplicity and lack of specificity as limiting its heuristic value (For a review of these issues, see Larouzee & Coze, 2020). For example, a practical issue is that links between different causal, temporal and individual factors or defence layers are not clarified by the SCM (Dekker, 2002; Luxhøj & Maurino, 2001) and the model does not in itself offer tools or guidance for practitioners to use in measuring and managing risks real world settings (Luxhøj & Kauffeld, 2003).

Against this context, Dawson and McCulloch's (2005) Fatigue risk trajectory (FRT), depicted in Figure 4, can be seen as a specific development of Reason's SCM that addresses some of these practical concerns for the hazard of sleep-related operator fatigue. It proposes how fatigue risks may incrementally build from latent organisational factors and conditions to affect operator performance 'on the day' using as its basis, evidence from the sleep scientific literature. Under the FRT, the chain of risks begins at Level 1 with the sleep opportunities afforded to the broader workforce by shift patterns, which influence the actual sleep obtained by individual operators (Level 2). According to the model, increases in fatigue risk exposure at these levels mean that operators are more likely to experience fatigue-related symptoms at work (FRT 3) that in turn, may develop into fatigue-related errors and incidents (FRT 4) which represent the proximal end of the risk trajectory prior to an accident.

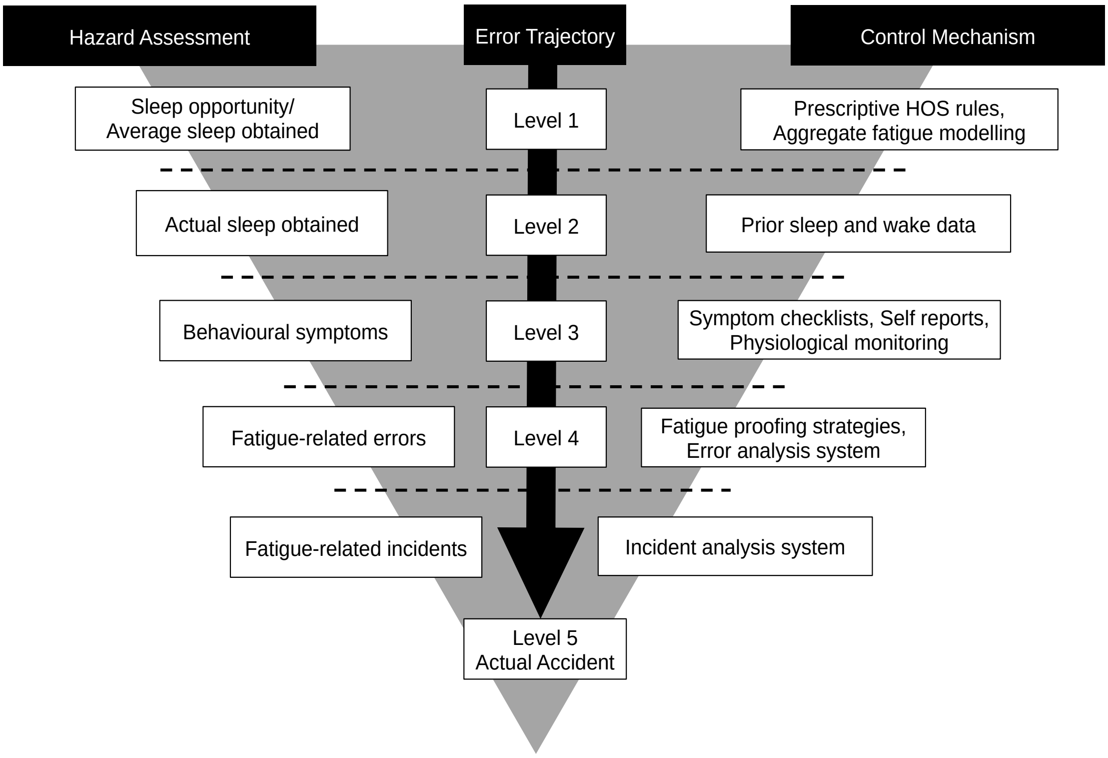


Figure 4. An illustration of Dawson & McCulloch's (2005) Fatigue-risk trajectory (FRT). A conceptual model outlining risk layers that precede a fatigue-related incident or accident, for which there are identifiable hazards and potential for control measures.

Phillips further developed the FRT by specifying organisational control and monitoring mechanisms that could be deployed as ‘defences’ at each FRT risk layer (Phillips et al., 2017). Against this framework, at Level 1, a practical starting place for the assessment of fatigue risk factors is the assessment of schedule-driven fatigue risk and quantification of whether shift patterns are likely to lead to insufficient sleep or to extended hours of continual wakefulness in operators during their duties. This is because duty hours dictate the length and circadian alignment of sleep opportunities available (Åkerstedt & Wright, 2009; Costa, 2015), and pre-duty sleep length and hours of continual wakefulness are crucial factors influencing on-duty sleepiness in all shift-work settings (Darwent et al., 2015); including aviation (Roach et al., 2012; Sallinen et al., 2017). At this level, Phillips and others (Dawson et al., 2012; Dawson & McCulloch, 2005) propose that the defence controls include aggregate fatigue modelling of rosters to identify elevated risks, and the existence of prescriptive rules which limit work hours. At FRT level 2, the risk is that the actual sleep obtained by employees may deviate from what was predicted or expected by the roster pattern. Hence the associated controls should establish the actual sleep and wake patterns achieved by the affected operators prior to and during their duty. At FRT level 3, the individual’s experience of fatigue and severity of associated symptoms whilst at work provides an indication of insufficient sleep achieved by employees (brought about by work-, home-life or other factors such as medical conditions). Hence fatigue risk at this level relates to the on-duty experience of fatigue-related symptoms that relate to neuro-behavioural fatigue-related decline and may impact safe operator performance and behaviours. This means that organisational controls which seek to monitor the experience of on-duty fatigue-related symptoms are needed and may take the form of passive physiological monitoring of aspects of fatigue or require active participation from employees through subjective self-reports against symptom checklists or subjective scales which assess aspects of neuro-behavioural decline. FRT levels 4 and 5 represent the levels on the trajectory where there has been performance errors or an operational impact to safety as a consequence of a fatigued operator. In safety practitioner parlance, control measures at this level can be thought of as important ‘lagging’ indicators (Manuele, 2009). Analysis of the frequency and types of fatigue-related errors made by operators provides an indication of the operational tasks which are less resilient or more prone to failure by fatigued operators, and provides an understanding of the factors and rates at which operator fatigue has developed to the point of threatening safe performance or indeed broader whole-system safety. Whilst predominantly developed and extended as a tool for exploring fatigue controls and countermeasures, the

FRT hence lends itself to being used as a framework for structuring research activities at each layer of fatigue risk. In this regard, one major implication of the model is the importance of investigating the fatigue risk exposure estimates across each of the different levels of the error trajectory and where possible, the relationships between these levels (Phillips et al., 2017).

It is the specificity of the FRT framework for guiding data collection and analysis on different layers of organisational risk that underpins its comparative superiority as a conceptual model both for measuring pilot fatigue and structuring research for this thesis. The FRT framework, as developed by Phillips et al. (2017), treats fatigue as a critical human factor limitation within the pilot safety-critical sub-system. It directs attention at the progressive accumulation of fatigue risks through successive temporal phases and the potential controls available at each phase to constrain pilot fatigue. Applied in this way, the FRT framework is not intended to be a comprehensive whole system accident prevention model, but rather is specifically designed for the analysis and measurement of precursor fatigue risks, and the fatigue-related break down of the human component, prior to a whole-system accident opportunity. The use of the FRT framework in this thesis draws on extensive findings from sleep science, as discussed in Chapter One. These findings regarding the biological processes of sleep-wake and circadian drives demonstrate that sleepiness accumulates over time and at a certain point manifests as a destabilising wake state with non-trivial performance decrements. The specific issues addressed in this thesis correspondingly require analysis of the trajectory of such sleep-driven fatigue risks and associated performance decrements across the pilot workforce, how accurately these are predicted by biomathematical modelling, how frequently pilots self-report increasing and severe sleepiness levels (as determined by laboratory validated scales), and how systematically such performance decrements and fatigue-related safety events are reported.

This framework differs in important respects from more recent theories and approaches to safety management, such as for example, Functional Resonance Analysis Method (FRAM) under *Safety II* (Hollnagel, 2018); Human and Organizational Performance (HOP; Conklin, 2012), Safety Differently (Dekker, 2015) and Systems Theoretic Process Analysis (STPA; Leveson, 2017). Such approaches challenge traditional theoretical models based on causality and argue the need to extend beyond a focus on identifying individual failure, especially of the human sub-system. Instead, their remit of interest is to understand and model the increasing interdependencies within complex socio-technical systems in order to achieve a comprehensive understanding and description of entire system reliability and interactive complexity. These approaches typically aim to describe operational activities and system interfaces within a map of system components and control processes, that is subsequently analysed to identify control weaknesses, in a process informed by employees and safety professionals (Provan et al., 2020). Such approaches do not provide precise evidence-based insights or scientific methods to appraise specific risk areas outside the domains of generic safety theorising and systems engineering. For example, approaches such as FRAM by design set the objective of not decomposing a system into its component parts but rather attempt to provide an overall understanding of how an organisation-wide socio-technical system works, or should work (Hollnagel, 2018). As such, the methods deployed include multiple investigative techniques involving interviews, observation, work audits, focus groups and surveys, to map critical processes (Patriarca et al., 2020) with the objective of improving system-wide safety and accident prevention rather than directly addressing quite specific, identified human component vulnerabilities. Other approaches that share this broader remit of interest also propose the need to focus on assessing ‘what goes right’ (Hollnagel, 2012) or the ‘presence of positive capacities’ (Dekker 2015), but in so doing, tend to describe theoretical approaches to safety learning regarding whole system safety over more selective practical methods to answer specific research questions that require highly specialised areas of analysis.

To answer the specific research questions in this thesis, the analysis of the human component vulnerability relating to fatigue requires at its foundation an understanding of the relevant sleep science and in particular an assessment of what constitutes an acceptable level beyond which pilots carry foreseeable safety risks, just as they would in the case of medical incapacitation or alcohol intoxication. Establishing and subsequently consistently measuring such a level beyond which fatigue confers safety-related performance decrements thus represents a crucial precondition for designing appropriate controls and constraints for maintaining human component reliability, and ensuring that fatigue does not compromise overall system safety. Understanding, measuring and managing pilot fatigue is a sufficiently complex area of study in its own right, which requires practical boundaries of analysis. The twofold challenge is that there is no established consensus on acceptable maximum levels of fatigue, and there is severely limited existing research on actual incidence and severity of in-flight sleepiness in UK commercial aviation and its potential operational impact. These are already identified problems on which pilots themselves have commented extensively, and for which there are separate relevant bodies of medical and sleep science which are directly applicable to the safety of the pilot sub-system. Thus, for the purpose of this thesis, there is need for a framework and methods that offer greater precision in the appraisal of fatigue risks across different organisational layers and points in time. As such, the research requirement is for a framework that can facilitate a quantitative approach based on sleep science in order to model both distal fatigue factors and measure in-flight real-time pilot experience, as well as structure the analysis of retrospective flight incident reporting. It is correspondingly the multi-level FRT framework that best meets this objective and sets practical boundaries to the required scope of analysis. In light of these considerations, the FRT framework has been selected for its decisive advantages over other conceptual approaches.

2.2 Aims and objectives

For the purpose of this thesis, the FRT has thus provided the conceptual background for the development of two studies that sought to contribute new evidence on distal and proximal fatigue risks in UK commercial aviation, to evaluate the contribution of such risks to unsafe operator performance and discuss the implications for policy and practice. These are briefly summarised below and depicted in Figure 5.

Study 1: Assessment of Aviation Safety Occurrence Reports

To appraise existing real-world safety data on fatigue risks in aviation using the FRT framework as guidance, Study 1 sought to produce a review of fatigue risk data derived from mandatory occurrence safety reports (MORs), the main safety occurrence data collected by the Civil Aviation Authority within UK commercial aviation. The aims were threefold; 1) to assess the quality and coverage of fatigue risk data (FRT levels 1-4) captured by the MOR database; 2) to investigate whether there were discernible associations between fatigue risk parameters and increased risks to pilot performance and emergent system safety and 3) to investigate whether there were emergent themes that characterised fatigue-related safety incidents within aviation. Against the framework, the retrieval of fatigue-related MORs also provided an indication of a safety occurrence rates for FRT level 4 over the last decade, to compare against other fatigue risk exposure levels in the FRT.

Study 2: Assessment of predicted and reported sleepiness and sleep events during normal operations

Study 2 sought to assess the predicted fatigue risk exposure of routine UK airline commercial pilot work schedules (FRT 1) and the incidence and severity of self-reported sleepiness (FRT 3) through a large scale field study of normal UK airline operations. The aims were 1) to provide new evidence on the aggregate fatigue risk exposure across the commercial aviation industry using bio-mathematical modelling, 2) to collect per-flight hour estimates of the incidence and severity of sleepiness experience by pilots across 18 UK airlines and 3) to trial a newly developed app to enable flight crew from different companies to self-report sleepiness symptoms and take part in research whilst at work.

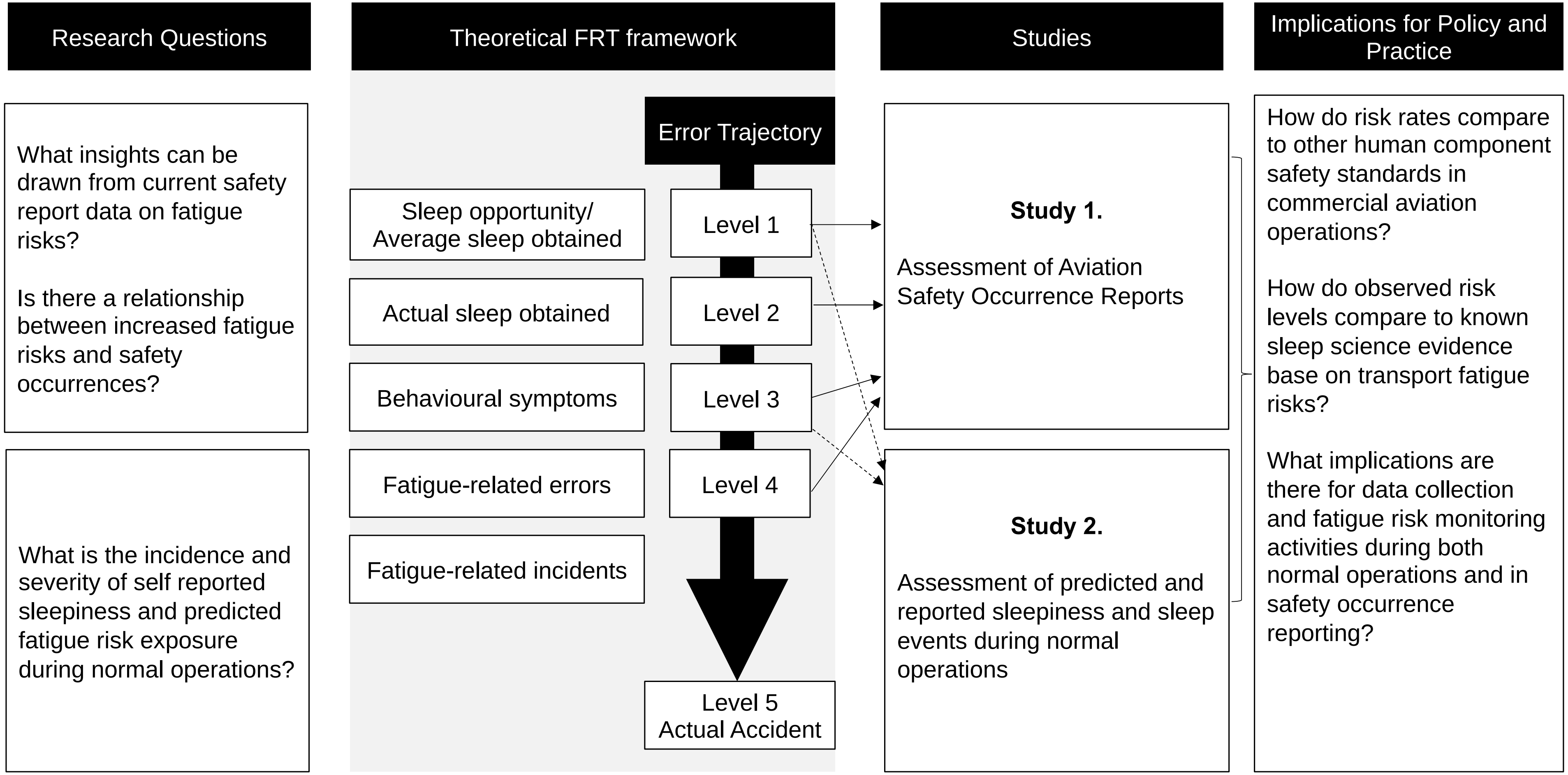


Figure 5. Summary of the research approach. The arrows indicate the FRT levels sought to be appraised in both Studies.

The objectives of Study 1 required a post-hoc assessment of fatigue-risk data in collected in MOR reports. Hence the study relied on assessing quantitative and qualitative data that had already been captured. However as a framework, the FRT was used as a guide for establishing the degree of coverage the presence or absence of work pattern data relating to sleep opportunity (FRT 1), actual sleep pattern achieved prior to the incident (FRT 2), self evaluated symptoms or use of validated scale (FRT 3), and performance-related errors (FRT 4). The fatigue risk metrics for this post-hoc study are described in full in Chapter 3, Table 5.

The objectives of Study 2 required an assessment of pilot work hours (FRT 1) and experienced sleepiness during normal duties (FRT 3), and hence required the direct collection and bio-mathematical modelling of work pattern and sleepiness data. For Study 2 the consideration of sleepiness measures are discussed in section 2.3 and bio-mathematical modelling in section 2.4.

2.3 Measurement of sleepiness

There are a number of ways to evaluate and quantify the phenomenon of sleepiness. As set out in the introduction to this thesis, the ‘sleepiness’ aspect of the operator fatigue construct has arguably received the most attention in safety-critical industries due to the enduring risk of sleep deprivation that arises from the combination of long working hours and rotating shift patterns which interfere with sleep drives (Dawson & McCulloch, 2005). Compared with other brain states covered under the umbrella term of ‘operator fatigue’ (Phillips, 2015), there is a greater degree of academic consensus on the definition, causes, measurement and associated performance decrements of sleepiness (Åkerstedt & Wright, 2009; Durmer & Dinges, 2005; van Dongen, Maislin, et al., 2003). For these reasons in scientific literature, ‘sleepiness’ is often considered the key phenomenon of interest with regards to the human factor’s term ‘operator fatigue’.

The state of sleepiness refers to the indistinct border between wakefulness and sleep (Ogilvie et al., 1989). Sleepiness measures vary in their sensitivity and specificity to fluctuations in alertness, which aspects of the sleep-wake transition they address, their suitability for use in different environments, and whether they are used for specific populations of people (for example the normal population versus patients with sleep complaints or medical conditions). For the latter category, there are a large number of bespoke measures suitable for assessing when sleepiness occurs either for atypical reasons or at inappropriate times unrelated to homeostatic or circadian drives. These instruments are often bespoke to particular patient groups, for example those with psychiatric conditions, medical conditions such as cancer, or those with primary sleep disorders such as narcolepsy, where the individual has a medical sleep complaint. Such measures were not considered since the participant group for Study 2 was a non-clinical population of licensed commercial airline pilots (certified as medically fit), and the objective was to assess state sleepiness levels across normal work hours. Aside from clinical measures, research instruments for measuring sleepiness tend to fall into one of two categories; 1) physiological (quantification of arousal decrease, physiological responses, latency to fall asleep when asked to do so) and 2) subjective (self-assessment of the subjective experience, self reported behaviours and sleep propensity) (Curcio et al., 2001; Shahid et al., 2010). The benefits and limitations of these methodologies are discussed below.

Physiological measures

Physiological measures of sleepiness include the measurement of brain activity, eye movements and eye-closure related measures, and autonomic indicators such as facial muscle tension, heart rate related measures and postural measures, among others (Doudou et al., 2020; Lal & Craig, 2001; Phillips, 2015). Sleep researchers in laboratory or clinical settings use techniques such as polysomnography to measure the electrical activity of subjects as they get progressively more sleepy and as they sleep (Kryger et al., 2010). Examples of these methods include the monitoring of brain activity using electrodes placed on the scalp (i.e., electroencephalograms or EEGs, or the recording of activity associated with eye movements (i.e., electro-oculograms or EOGs). These methods are usually considered to be the most valid standard measures for assessing physiological sleepiness because they directly measure nerve cells firing at different frequencies (Balandong et al., 2018; Phillips, 2015; Roehrs & Roth, 2001). At a general level, in sleep-deprived subjects, there are increases in alpha activity (8-12Hz) and theta activity (4-7Hz) which indicate increased sleepiness. Alpha activity (8-12Hz) is thought to characterise a relaxed wake state and theta activity (4-5Hz) is associated with decreased processing activity and recognised as the first stage of sleep (Liu et al., 2009; Roehrs & Roth, 2001). Humans may technically have 2-3 minutes of EEG defined sleep before the majority of subjects are actually aware that they had been sleeping when they wake up (Prabhu, 2021; Rosekind et al., 1994). This despite the fact that most experience precursory feelings of sleepiness prior to sleep (Reyner & Horne, 1998). EEG signals may show a progression in alpha and theta activity from wakefulness to stages 1 and 2 sleep, or sleep may be preceded by microsleeps, which are represented by 5 or more seconds of alpha dropout and an increase in theta activity (Harrison & Horne, 1996). Thus EEG can detect brief periods of sleep even if these occur outside of the subject's awareness. The sensitivity of EEG measures hence confers a major methodological benefit for being able to identify the stages of the wake state beginning to de-stabilise and transition into sleep. In addition, the passive aspect of this monitoring technique means that subjects (in this case pilots) do not have to actively and consciously self-report and provide sleepiness data. Self-reporting by contrast involves an evaluation process that by definition will not identify sleepiness or sleep episodes outside the subject's awareness. However, most work involving EEG measurement is generally conducted in a laboratory setting, where both the surrounding environment and the subjects' behaviour is under stricter control. In these settings, the assessment of a subject's latency to fall asleep is one of the most widely used measures of sleepiness.

The use of EEG in Study 2 was considered as some authors working in the transport field have been able to bring EEG/EOG methods outside of the laboratory for the simple passive recording of sleepiness via compact portable devices such as Medilog tape recorders, which are designed for continuous EEG monitoring via small battery powered devices (Rosekind et al., 1994). These devices permit monitoring of the EEG from selected brain areas via electrodes fixed to the scalp. Researchers using devices like these have been able to assess sleepiness among long-haul drivers (Kecklund & Åkerstedt, 1993), train drivers (Torsvall & Åkerstedt, 1987) and healthy, sleep deprived drivers in simulated driving protocols (Balandong et al., 2018; Reyner & Horne, 1998). These latter studies showed that EEG measures collected via these devices correlate reasonably well with both subjective experience of sleepiness, and with deteriorating performance in healthy sleep deprived drivers. On rare occasions, EEG has also been used in the commercial aviation setting, providing some precedent for its use in research (e.g. Cabon et al., 1993; Wright & McGown, 2001; Rosekind et al., 1994). However, the main disadvantage of techniques based on biomedical signals is that they are often less practical to use outside the world of academia and heavily constrain the sample size and location. Such recording devices invariably require multiple sensors and cables to be attached carefully on the body, are expensive methodologies, and require specialist technicians to be involved in their set up with subjects. Such devices also need to be looked after, cleaned and reset properly for each individual. Thus any continuous monitoring of sleepiness via this method is likely to depend a great deal on whether researchers have available resources and the environmental setting not presenting a problem for their use. The use of such devices during real world operations, such as aviation commercial operations furthermore necessarily depends on both the airline operators' risk assessments over the likely hazard of using such equipment during flight (e.g. taking into account operator distraction, device manoeuvrability around other functional equipment such as headsets, etc) alongside agreements from all of the flight crew (not just subjects) on a given operating flight. For these practical reasons use of EEG for Study 2 was not favoured. A more specific set of concerns for the use of EEG in real world settings or with active participants is that of ensuring good signal quality, and separating the true sleep-driven frequencies from other noise, movement or cognitive artefacts. For example, researchers using EEG in the driving simulator context have warned how cognitive task demands may act to confound or alter classical signs of sleepiness, making them less easy to detect (Horne & Reyner, 1995b). Thus while EEG measurements are sensitive to sleepiness, recordings typically require substantial off line processing and may further require detailed experience and expertise on behalf of the researcher on the analysis of signal recordings in the environment in question. Taken together for the aviation field operations, such measures were considered to be less appropriate for wide scale use. However, their use in simulator flights may be both possible and, where coupled with specific experimental designs, are likely to provide the most detailed insights on the physiological sleepiness states of subjects.

Oculomotor Activity

Various ocular variables (including duration of blinks, incidence of long eyelid closures, relative velocity of blinks and saccades) have been shown to provide relatively good objective quantifications of levels of sleepiness (Abe et al., 2011; Åkerstedt et al., 2010; Johns et al., 2007; Komada et al., 2013). In particular, blinks associated with the onset of micro-sleeps or sleep appear to be characterised by unique properties (Ftouni et al., 2013). Compared to well-rested individuals who typically show closing and reopening phases of blinks at around 150 and 100ms in duration respectively (Stern et al., 1984), eye blink rates in sleep-deprived individuals are typically longer, with the closing phase exceeding 250ms, and the reopening phase lasting between 100-150ms. In addition, eye closure during the closed blink phase is typically extended, between 250ms to several seconds (Ftouni et al., 2013; Stern et al., 1984). Periods of eye closure lasting more than 500 ms have been termed 'slow eyelid closures' (Alvaro et al., 2016). Alterations to the frequency, amplitude and duration of blinks have been found to occur prior to sleep initiation as measured by EEG and EOG techniques in sleepy subjects under a wide variety of conditions (Åkerstedt et al., 2014; Alvaro et al., 2016). Thus, these ocular parameters, and a number of other composite variables derived these ocular parameters are thought of as proxy indicators of sleepiness or 'drowsiness' (Jackson et al., 2016; Santamaria & Chiappa, 1987). For example, percentage of eyelid closure over the pupil over time (PERCLOS) alongside EEG-derived correlates of sleepiness have been used to track sleepiness in the driving setting (Åkerstedt et al., 2005; Trutschel et al., 2011), and have been shown to predict driving performance. For example. with increased blink duration, studies have found an increased crash rate and variability in lane position (Åkerstedt et al., 2005; Anund et al., 2009). Furthermore increases in blink duration, amplitude variability (Lal & Craig, 2001) and measures such as PERCLOS have been associated with deterioration in driving performance (Liu et al., 2009).

One of the main difficulties associated with the measurement of ocular parameters in research relates to equipment constraints and their practical implementation. Many technologies which use eye-tracking and video-based devices to measure ocular parameters are difficult to implement outside laboratory conditions, and may be affected by variable environmental light conditions (Johns et al., 2007). In recent years, the desire to measure 'operator drowsiness levels' has yielded the commercial production of a number of portable measurement devices that purport to track a number of ocular variables. One commercially available system, Optalert^{TM}, that has been used in a variety of applied field contexts, uses infrared (IR) reflectance oculography with infrared (IR) transducers (LED emitters and a phototransistor receiver) attached to a glasses frame that monitors eye and eyelid movements continuously and relatively unobtrusively (Johns et al., 2007). Each pair of frames are tailored to the individual's face, such that the transmitter receiver rests at the correct angle to track the upper eyelid. The variables tracked by these frames include the standard deviation of the duration of blinks, the mean duration of eyelid closures, the relative velocity of eyelid closing movements during blinks, and the amplitude velocity ratios of eyelid reopening movements during blinks, all measured per minute. A weighted combination of these ocular parameters provides a single measure of alterations in drowsiness levels, which the developers operationalised as the John's Drowsiness Scale (Johns et al., 2008). The practical attractiveness of this technology for use in research on operator drowsiness is that it satisfies the following inherent features of the operational environment: it is functional in low levels of light, is non-invasive, provides minute-to-minute detection, can theoretically accommodate individual variability, and is housed in a portable device. Some studies have suggested that it is validated against objective performance variables such as PVT lapses and subjective sleepiness (KSS) (Jackson, Kennedy, et al., 2016; Johns et al., 2008), although it is clear further work needs to be conducted to confirm this relationship. For example, a recent flight simulation study assessing a variety of ‘fatigue’ measurement techniques suggested Optalert, amongst others carried a significant amount of noise and data losses during operation, to the point where Optalert data were not able to be fully presented (Thomas et al., 2015). As part of the background preparation for this thesis, the Optalert equipment was tested with a UK air ambulance operator, and various issues with the design of the frames and eyelid-movement data were noted. Two key issues in the data were that of signal dropout (large gaps in the eyelid-movement data) and mis-attribution of eyelid blink data by the software when the pilots were doing tasks that shifted their gaze downwards (for example periods where the pilot was reviewing checklists) which provided a signal of long eye-closures indicative of micro-sleeps or longer sleep periods. Aside from their expense, many drowsiness monitoring technologies such as Optalert appear to still be in the prototype stage of development, and thus their utility in different types of operational environment is still yet to be established.

Whilst in general, physiological measures offer benefits in the assessment of sleepiness in individuals, particularly in terms of sensitivity, they are not immune to various confounding effects of both external factors (eg environmental glare, humidity), equipment failure or situational/ person specific factors (eg. non-sleep factors affecting heart rate or eye blink rate, or behavioural factors out of the researcher’s control). However, for Study 2, the common methodological limitation for most physiological measures available to assess sleepiness is that they are not typically suitable or feasible for large scale studies seeking greater participant numbers, as most require use of either expensive measurement devices, or devices that are individually tailored to participants.

Subjective measures

Subjective measures of sleepiness measure how sleepiness or ‘operator fatigue’ are experienced by the individual and mainly rely on rating scales and questionnaires (Shahid et al., 2010). At a general level, subjective self-report measures are relatively quick and easy to self-administer, inexpensive, reproducible and do not require any prior training. Within the sleep science literature, subjective rating measures tend to fall into two categories assessing ‘trait’ (longer term, general daytime sleepiness levels) or ‘state’ sleepiness (short term variations in alertness level). Trait scales typically have been used in clinical settings to establish the presence of pathology via likely sleep onset occurring at inappropriate times. For example, the Sleep Wake Activity Inventory (SWAI, Rosenthal et al., 1993), designed to assess different components of sleepiness and distinguish normal from pathological sleepiness, contains fifty-nine items with an accompanying 1-9 Likert-type scale used for each item, denoting presence or absence of a behaviour over the last seven days. For assessment of excessive daytime sleepiness, the SWAI appears to show good internal consistency (chronbach's $\alpha = 0.88$), and has been validated using standard multiple sleep latency tests (physiological assessments of time taken to fall asleep). Some shorter self-rating questionnaires such as the Epworth Sleepiness Scale (Johns, 1991) can be used in both clinical and research settings. The ESS is an eight item instrument which measures daytime sleepiness through assessing the participant's overall tendency to fall asleep in different situations typically encountered in daily life (eg. 'sitting and reading') across several weeks through a scale of 0-3 for each situation, the sum of which produces the participant's score. Although the ESS is discriminative (insofar as it provides a cut off score of ten and above indicating abnormal or pathological sleepiness) it is not bespoke to a patient population and is one of the most widely used daytime sleepiness measures for general research (Shahid et al., 2010). The ESS has shown good test-retest reliability ($r = 0.82$) and internal consistency ($\alpha = 0.88$) (Shahid et al., 2010), however recently the cultural applicability and ambiguity of certain items have been questioned by some researchers, suggesting that some of the items may be either dated or not applicable to different population groups (Alqurashi et al., 2021). Whilst trait measures of sleepiness such as the SWAI, ESS and revised versions of these inventories have become established within sleep science literatures, they are not valid measures for assessing hourly variations of alertness across a work shift, and as such were not considered suitable for the objectives of Study 2.

The 'state sleepiness' category of subjective measures includes self-evaluations which are used to assess short term changes in sleepiness through scales such as the Stanford Sleepiness Scale (SSS, Hoddes et al., 1972; Mitler et al., 2005), Karolinska Sleepiness Scale (Åkerstedt & Gillberg, 1990), and Visual Analogue Scales (Dittner et al., 2004). Such measures seek to assess an individual's fluctuations in sleepiness across their waking hours, and as such, repeated subjective ratings can be made by the individual at multiple time points. At a general level, instruments seeking to assess within-subjected repeated measures should not be unduly burdensome to the participant, as this may influence the participant's responses in a variety of ways (Monk, 1989). Brief single scale measures are therefore more appropriate for determining short term changes in sleepiness in response to sleep, circadian factors as well as environmental stimuli and drug effects (Shen et al., 2006).

Visual analogue scales (VAS) can be used as markers of state sleepiness where subjects are asked to indicate on a 100 mm line their subjective experience between two extremes of alertness “very alert” and sleepiness “very sleepy” (Monk, 1987). The numerical value derived from the test is the number of millimetres measured from one end of the scale to the line bisection made by the subject, that corresponds to their perceived level of alertness / sleepiness on a continuum between the two extremes. Implicit to the design of such a scale is that it is continuous, differentiating it from discrete stages used in item description scales. Hence, a useful property of VASs is that the construct of what is being measured is intuitively clear. VASs provide a visual illustration of a line denoting increased or decreased magnitude (in this case between alertness and sleepiness) which exists on a single continuous dimension, without complex words or phrases more open to interpretation, and avoid the need for participants to have to ascribe sleepiness sensations to a verbal category that may be ambiguous or lack cultural applicability. Further, VASs allow for greater sensitivity and variation between participants, and for analysis purposes the continuous scale allows for its use in parametric statistics. Some limitations of VASs however include the need to deploy fine visual and motor skills in order for the measurement to be filled out, and some authors have suggested that participants feel hesitant in using the extreme ends of 100-mm lines (Shahid et al., 2010). The main practical drawback of VASs however occurs at the data entry and analysis stages. Unless electronic forms are used or there are programming options for the measurement of the lines to be performed in an automated manner, the input of data can be extremely time consuming, particularly with large participant numbers and repeated measurements across the study period.

Another widely used sleepiness scale is the Stanford Sleepiness Scale (SSS; Hoddes et al., 1972), typically but not exclusively used for quantifying state sleepiness levels in patients with sleep problems. It is a Likert-type scale with seven ranked levels, which requires participants to select the level that best describes their current state at the time that the assessment is administered. The scale ranges from ‘1- feeling active and vital, alert, wide awake’ to ‘7-almost in reveries; sleep onset soon; lost struggle to remain awake’. Some studies suggest the good concurrent validity with physiological sleepiness because the SSS appears sensitive to sleep deprivation. For example, repeated measurement using the SSS has been found to correlate with sleep latency tests and sleep restriction in normal subjects, and the scale developers have suggested that ratings made every 15 minutes are sensitive to fluctuations in sleepiness (Shen et al., 2006). For conceptual clarity on what is being measured, single dimension scales should ideally use the same parameter with an ordered increasing magnitude. In this regard, while the SSS has been extensively used, its items contain a mixture of adjectives indexing sleep propensity, energy/fatigue and cognitive performance, and factor analyses have suggested a 3 component structure along similar lines (‘alertness/sleepiness’; ‘loss of control’, and a ‘cognitive factor’) (Shen et al., 2006). Thus the SSS does not appear to directly measure sleep propensity, but rather a more global measure of sleepiness. These different aspects of subjective experience, whilst important, do not logically fit in a single dimension ordered scale. Methodologically this means that the interpretation of high SSS scores is less certain, or potentially problematic. This particular limitation of mixing parameters within a single dimension ordered scale can be extended out to a number of other subjective scales used to assess ‘operator fatigue’ in applied fields, including the Samn-Perelli scale (SP, Samn & Perelli, 1982), frequently used in commercial aviation (e.g. Gander et al., 2013, 2015; Honn et al., 2016). This problem of the scientific interpretation of mixed parameters on a single dimension scale is compounded where scales contain ambiguous terms which bear unclear relationships to one another. For example, the SP scale contains descriptions such as ‘Very lively, responsive but not at peak’ and ‘Okay, somewhat fresh’.

The state scale that appears to provide many advantages for use in Study 2 is the Karolinska Sleepiness Scale (KSS, Åkerstedt & Gillberg, 1990). The Karolinska Sleepiness Scale KSS is a nine-point one dimensional subjective scale that only indexes the alertness-sleepiness continuum. Its descriptions at each level on the scale (see Table 3 below) are clear insofar as they unequivocally relate to alertness and sleepiness and at face value, follow an intuitive increasing ordered magnitude in severity of sleepiness with increasing scores on the scale. Rating on this scale can be done quickly, repeatedly and provides the individual with a simple measure of their sleepiness that reflects the psycho-physical state experienced in the last five minutes.

Table 3. The Karolinska Sleepiness Scale (Kaida et al., 2006; Kecklund and Åkerstedt, 1993; Sagaspe et al., 2008)

1	Extremely alert
2	Very alert
3	Alert
4	Rather alert
5	Neither alert nor sleepy
6	Some signs of sleepiness
7	Sleepy but no effort to keep awake
8	Sleepy, some effort to keep awake
9	Very sleepy, great effort to keep awake, fighting sleep

The KSS is considered a reliable tool for use both within laboratory and field research for evaluating changing levels of sleepiness (Åkerstedt et al., 2014) as it is both sensitive to sleepiness fluctuations (Åkerstedt et al., 2014; Åkerstedt & Gillberg, 1990), and validated against performance and EEG variables (Kaida et al., 2006b; Kecklund & Åkerstedt, 1993; Sagaspe et al., 2008). Importantly for construct validity of the KSS measure, the KSS has been shown to be sensitive to known sleep drivers; for example, KSS scores rise with increased periods of sustained wakefulness (Åkerstedt & Gillberg, 1990; Åkerstedt et al. 2014) and strongly correlate with time-of-day (Kecklund & Åkerstedt, 1993).

In terms of concurrent validity, KSS scores have also been shown to correlate with EEG measures of sleepiness (alpha/theta activity) (Åkerstedt et al., 2014), which aids interpretation of scores particularly for the upper end of the scale. Indeed, Åkerstedt and colleagues have shown that the first signs of physiological sleepiness start to occur at KSS level 7, are accompanied by the gradual increase of heavy eyelids and at levels above 7, and there is a marked increase in EEG-related sleep intrusions, long eye lid closures at KSS 8, suggesting a point at which top-down attempts to stave off sleep are breaking down (Åkerstedt et al., 2014). At KSS 9, EEG-identified sleep intrusions and other ocular parameters are pronounced, reflecting a de-stabilising wake state in the participant.

Of further validation interest is the finding that the KSS appears to correlate reasonably well with behavioural performance decline. For example, a number of studies have demonstrated midsized to strong correlations between KSS ratings and reaction time/vigilance performance (mean $r > 0.65$) (Kecklund et al., 1994) or lapses ($r = 0.57$) (Kaida et al., 2006b) during a night awake in the laboratory, and with performance decrements (lane crossings) during night driving (Sagaspe et al., 2008). It seems that in particular, high ratings on the KSS scale (KSS 8 and 9) show a close relation with physiological indicators and performance indicators in driving (Anund et al., 2009; Ingre et al., 2006; Reyner & Horne, 1998). For example, KSS levels of 8-9 appear to be related to not only marked increases in sleep intrusions in the EEG and EOG measures but also in driving crash risk (Åkerstedt et al., 2014; Sandberg et al., 2011).

Thus, the attractiveness of the KSS scale is that, in comparison to most other subjective scales, it has received scientific validation against objective physiological measures of sleepiness. It can be administered easily and quickly, and shows sensitivity to sleep homeostatic and circadian parameters known to affect physiological tiredness. In a review of the work conducted using the KSS measure, Åkerstedt et al. (2014) highlighted a few inconsistencies with respect to the sensitivity of the scale. The main finding was that in studies using partial sleep deprivation, the increase in KSS rated sleepiness was much milder than the decline in PVT performance (Van Dongen et al., 2003). The authors suggest that pre-rating or situational conditions may have accounted for some of this discrepancy. However it is also possible that this dissociation between subjective experience and performance is stronger in some individuals. As a subjective measure it clearly shares the same vulnerability to conscious or unconscious bias as other self-report scales.

Logically it could be argued that VASs provide a more accurate reflection of the continuum between alertness and sleepiness experienced by individuals. However, there are different perspectives on this and both KSS and VASs have their advantages. For example, whilst some highlight the increased sensitivity that VASs offer, others have suggested that in practice, individuals better understand the conceptualisation of a scale and are better assisted by verbal descriptors placed at even intervals on the line in between two extremes (Wewers & Lowe, 1990). The manual element of VAS rating furthermore should not be underestimated as a potential confounding variable, where this is completed remotely on, for example a smart phone in a cockpit environment where factors such as turbulence may be present. This may introduce noise and coarseness to the data not seen in traditional paper and pen formats.

Given the mixture of benefits and burdens associated with both objective and subjective methodologies, ideally most research initiatives should seek to combine both types of measure. An additional benefit of the KSS scale relating to the research aims of Study 2 is that it is also used as an output score in a number of biomathematical models. Hence, for efforts that seek to both predict and measure fatigue, the KSS helps to provide consistency in terms of what is being measured, and hence allows more valid comparisons to be made. Against this context, the KSS was selected for use in Study 2.

2.4 Biomathematical modelling of fatigue risk

The use of biomathematical models to predict sleep opportunities and on-duty sleepiness has become particularly important in work environments where operator fatigue risks are elevated by the intrinsic nature of different types of shift patterns. At a general level, such models seek to quantify the effects of sleep homeostasis and circadian processes on the regulation of alertness and performance, and in so doing, predict the timing and magnitude of sleepiness related responses, and consequent ‘fatigue related risk’ to performance, particularly in shift-work contexts involving sustained hours of work, rotating duty starts or transmeridian travel (Mallis et al., 2004). The main variables which models seek to provide parameters for are sleep length, circadian rhythm and time awake, since sleepiness is mainly regulated by these three processes. The two main inputs to a biomathematical model to provide the relevant timing information are work-rest schedule or sleep data (obtained from for example a personal sleep log or wrist-movement technology such as an actiwatch).

The most widely accepted underlying model for predicting sleepiness and the likely sleep-wake cycle in humans is the two-process model, that mathematically charts the naturally occurring interaction of sleep and circadian factors (Borbély, 1982; Borbély & Achermann, 1992). This model (the equations of which are publicly available) was developed from sleep deprivation study data across a number of experiments using EEG, and its original purpose was to make predictions regarding the timing and duration of sleep (Borbély, 1982; Dawson, Noy, et al., 2011). Given a specific pattern of work, the two-process model is capable of making predictions regarding the timing and duration of sleep that an average person would experience, to a reasonable degree of accuracy (Dorrian et al., 2012). Subsequently the two-process model was extended to make predictions of alertness and sleepiness during the wake state and a further process W (Waking) was added as an option to capture the initial transient state of lower arousal immediately after awakening (Dawson, Noy, et al., 2011).

The homeostatic sleep drive (endogenous pressure to sleep or wake) is modelled as a pattern of declining alertness with increasing periods of continual wakefulness, and the recovery of alertness during sleep. The circadian processes (endogenous diurnal variation in alertness) are described by the interaction between the circadian rhythm lasting twenty-four hours and the ultradian rhythm lasting twelve hours. The aggregation of these processes produces an estimated level of sleepiness at any given moment (CASA, 2014);

- Process S (Sleep) refers to homeostatic pressure. Sleep onset occurs when process S reaches a high threshold level and wake-up occurs when S drops below a low threshold level. During sleep, process S decreases in an exponential fashion. Conversely, lack of sleep and/or extended duty time directly increases process S.
- Process C (for Clock/Circadian) is a sinusoidal function that schedules sleep to occur during night time and to cease during the daytime over a period of around 24 hours. This is influenced by environmental factors such as light availability. This is an independent process from Process S which takes into account time awake. Certain conditions relevant to aviation such as rapid time zone transitions and variable shift patterns directly impact on Process C.

The overall effect of these processes is that where normal sleep has been obtained, humans are alert throughout most of the day, apart from a small dip in the early afternoon, and then this alertness decreases quite rapidly as the night progresses, where the drive from both sleep and circadian factors is towards sleep (Basner, et al., 2013). As described in Chapter 1.2, most, if not all established commercial vendors who provide models of sleepiness that are integrated into a usable computer application, appear to be founded on this baseline formulation (CASA, 2014).

One of the major difficulties in assessing differences between bio-mathematical fatigue models currently is that there is commercial interest and sensitivity around the tweaks made to the basic underlying two-process formulation. This means that regrettably, for researchers with interests in open scientific enquiry, in general there is a reduced ability to scientifically assess reported claims of the model capabilities and compare properties of these models, and it would not be financially possible to gain licenses for all available bio-mathematical models. Set against this context, one important comparison paper of seven major bio-mathematical models of ‘human fatigue’ was published by Mallis and colleagues, who conducted a set of workshops in the early 2000s (Mallis et al., 2004). In their review Mallis’s team invited the developers of seven commonly cited bio-mathematical models (Two-process model, Sleep-wake predictor; SAFE model; Interactive Neurobehavioural Model; SAFTE model; FAID model and CAS Model) to complete a survey of their capabilities, inputs and outputs in order to produce a summary framework for comparing such models. At that time, Mallis concluded that all models reviewed had been ‘fundamentally influenced by the two-process model of sleep regulation by Borbély’ but that there was ‘…diversity in the number and type of input and output variables, and their stated goals and capabilities’ (Mallis et al., 2004; p1). It should be noted that a similar regulatory impact assessment paper by the Australian Civil Aviation Authority came to the same conclusion a decade later (CASA, 2014). From both of these reviews, it appears that all maintained the circadian factor of the two-process model as a key component, but the main ways in which the models differed was with respect to the role of sleep and work times as input factors for prediction. For example, Mallis found that four of the seven models had work time as their sole input variable, whilst the others relied on retrieval of sleep timing for their model input. In terms of outputs, models also varied in terms of their prediction goals and focus. In this regard, at the time, five of the models sought to predict results from laboratory experiments, field, and operational data (Two-process model, Sleep-Wake Predictor Model, FAID Interactive Neurobehavioural model and SAFTE), while three models (CAS, FAID, SAFE) were developed with less focus on predicting laboratory experimental and field results, in favour of different validation contexts.

Although perspectives vary on the relative utility of these goals and the development of validation in different contexts, from the academic perspective of this thesis, a priority was to use a model for which the predictions were based on known published evidence in sleep and circadian-mediated physiology, and objectively obtained performance outcomes. There are limits to this approach, insofar as it is acknowledged that all models require a focus on continued validation in different operational contexts, and some models not based on prior experimental data may have greater context specific validation data. However, implicit in the design of any of these models is that their underlying algorithms that estimate sleep closely approximate actual sleep achieved by operators during their shifts (Darwent et al., 2010). In this regard, model vendors that have published data to support their validation and lay bare any limitations in their data are favourable for research purposes. The Sleep-Wake predictor model (SWP) fares well in this regard. In its original version the SWP was used to predict the sleep patterns of pilots on long haul trans-meridian operations (crossing multiple time zones) (Darwent et al., 2010). The model was seen to generate average estimates of sleep timing of pilots with good predictive power for long haul aviation settings but in early development stages, it underestimated the amount of recovery sleep achieved by pilots during their rest periods (Hursh & van Dongen, 2011). However, continued development of this model in shift-work settings led to its improvement in sleep length predictions and additional validation in other non-transmeridian schedule patterns (Åkerstedt, 2008; Darwent et al., 2012). For example, Darwent et al. (2012) conducted a study using this model with sample of 225 train drivers who collected work/rest and sleep/wake data for two weeks during normal operations. The authors found that observed and predicted sleep periods showed high levels of agreement of 85%, which by any human modelling standards represents a robust degree of accuracy. This validation is important, since biological principles and laboratory data alone are not sufficient to explain the range of sleep behaviours exhibited in shift work settings.

Against this background of different considerations, the Sleep-Wake predictor (SWP) model has many advantages for the objectives of Study 2, and thus was selected. The SWP is a ‘pure’ implementation of the 2-process model, as it predicts likelihood of sleep onset and sleep termination based on physiological parameters. However, it also predicts alertness from sleep / wake patterns or from work pattern timing data, which means that pilot work schedules can be used as the input data, without the requirement for widescale collection of sleep logs. It was created using mainly group results from subjective alertness data collected in experiments on altered sleep/wake patterns, and subsequently validated against performance in driving, aviation and other shift work settings (Åkerstedt, Kecklund, et al., 2008; Åkerstedt & Folkard, 1996; Gundel et al., 2007; Ingre et al., 2014). Reflecting Borbély’s 2-process model, the SWP assumes an exponential fall of alertness during continuous wake hours, and an exponential rise of alertness during sleep, a circadian rhythm of alertness with a peak at 16:48. It also includes the option of an exponential sleep inertia factor just on waking. No further assumptions are made other than that of a normal 8-h sleep when starting the simulation. The model has been validated in many studies of shift work, mainly as group results, and it has been used to account for long and short sleep periods. While the software interface is basic, the SWP developers have been more transparent than other model vendors on the science underpinning the mathematics of the model itself, and in some major ways is well suited for research purposes. First, the authors have published the full equations used in the model in the public peer-reviewed domain (Åkerstedt et al., 2008) and are available for use without charge, although an interactive version can be purchased. Second, the SWP permits the extraction of raw output data (e.g. start and end times of all predicted sleep periods) from the model, whereas other models appear more suited to providing summary ranges or pre-installed graphics (e.g. low risk, medium risk or high risk for operator fatigue) for practitioner risk matrices. Third, the SWP provides an output prediction of subjective alertness using the KSS, which aids the interpretation of prediction values. The main output data from the SWP are the predicted sleep times associated with work pattern data and associated the predicted alertness curve across time, in the form of the 1-21 point generic scale or the Karolinska Sleepiness Scale (KSS 1- 9).

Chapter 3. Coverage of fatigue risk data in Mandatory Occurrence Reports

At a regulatory level, for many years the belief that monitoring near misses can improve safety has been embraced by aviation industry. In UK and European civil aviation operations, safety occurrence reporting is governed by European Regulation 376/2014, which views the reporting, analysis and follow up of safety occurrences to be vital set of activities in a proactive and evidence-based safety system. The mandatory reportable occurrence (MOR) database contains the safety reports submitted to the UK Civil Aviation Authority, for which the Commission Implementing Regulation (EU) 2015/1018 of 29 June 2015 provides a list classifying occurrences in civil aviation to be mandatorily reported (Commission Implementing Regulation (EU) 2015/2018, 2015). Such MORs represent occurrences, or pre-defined safety margin breaches which are legally mandatory to report, since they are perceived to be important indicators of risk. Pilot fatigue represents a mandatory reportable occurrence, covered under Annex III (3) articles 6 and 7;

(6) Fatigue impacting or potentially impacting the ability to perform safely the air navigation or air traffic duties.

(7) Any occurrence where the human performance has directly contributed to or could have contributed to an accident or a serious incident.

One of the clear challenges associated with this requirement is that there is no industry consensus on fatigue reporting measures to afford substantial consistency in the way in which this reporting regulation is interpreted, and often severe in-flight fatigue goes unreported. Hence investigations pertaining to the safety impact of distal and proximal fatigue risks require an analysis of existing report data to understand the reported rates of fatigue-related incidents and whether fatigue risk factors may be appraised from the report data.

The purpose of Study 1 was to assess the current coverage and quality of fatigue risk data included in a 10% random sample of UK fatigue-related MORs, with reference to the hazard levels described in Dawson & McCulloch’s (2005) fatigue risk trajectory framework. In tandem, Study 1 set out to synthesize relevant themes arising from the report narratives and highlight reported behavioural symptoms and fatigue-related errors and outcomes recorded in the MOR reports.

3.1 Ethics Statement

The purpose of Study 1 was to better understand serious reportable occurrences associated with pilot fatigue within commercial aviation operations. The report data constitute real world narratives of occurrences that may have compromised flight safety, and as such these data are not only commercially sensitive, but further, are associated with individual pilots who have submitted safety reports in good faith. Although the reports are available to interested responsible parties through the form of a freedom of information request, this must be for activities allied to the purpose of the advancement of flight safety. Hence, to protect the interests of an open, honest and protected reporting culture to the CAA, it was important that during the course of this research, no personally identifiable details from the report narratives were disclosed, and no airlines were mentioned. The reporting data were hence planned to be disidentified and redacted appropriately, in line with this aim. As a pre-requisite to the transference of sensitive safety data, the database entries provided by the CAA were disidentified at source, and as such the names and most personal information had been omitted. This measure notwithstanding, arrival and destination locations and aircraft types when paired together provide enough information to identify the airline of the reporter. As such, once the data had been forwarded from the CAA, the locations and aircraft types if included within the report narratives were redacted, and excluded from the analysis. The purpose of the exercise was not to expose a series of events that could be related to single safety reports, but rather to examine fatigue risk data and conduct a content analysis of any themes across a sample of reports drawn from the dataset.

3.2 Study Procedure

Figure 6 sets out the research stages taken in this study and associated report numbers. Further details of each stage are described below.

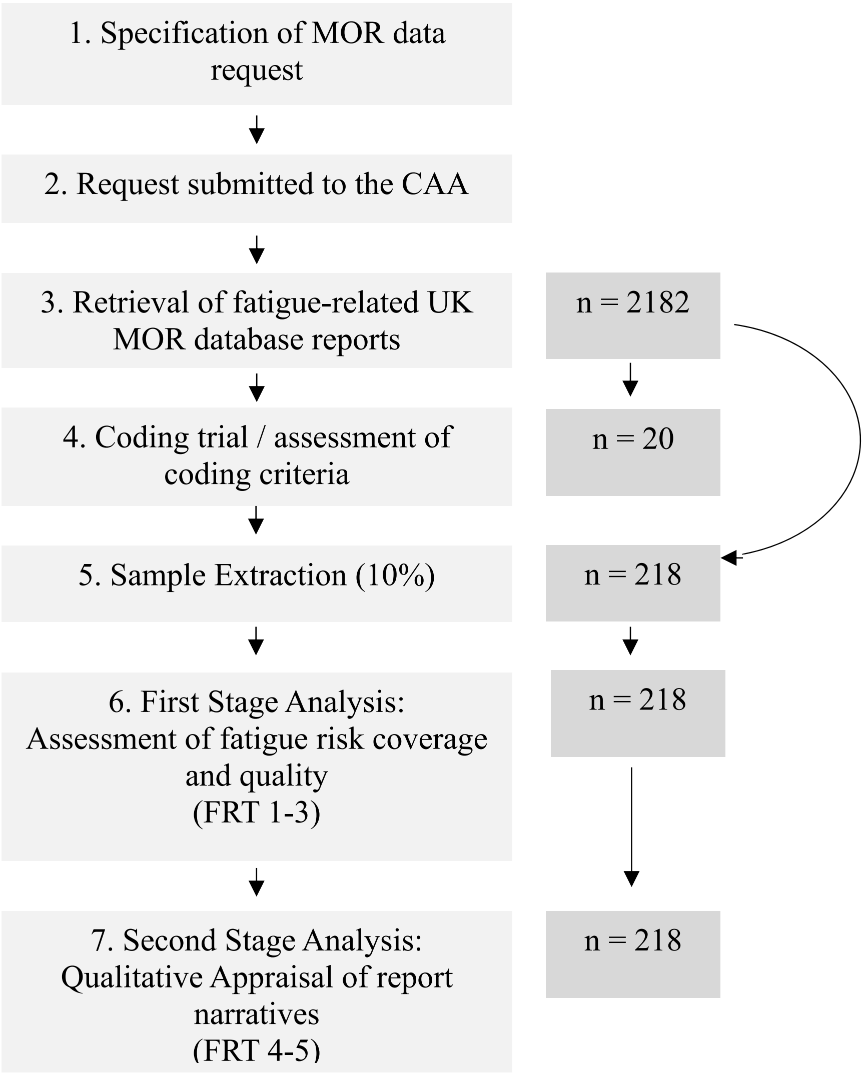


Figure 6. Research Process stages undertaken in Study 1. Number (n) of mandatory occurrence reports (MORs) assessed are set out against each stage *(right)*.

Stage 1. Specification of the MOR Data request

First a SRG1605 application form for data release was retrieved from the CAA MOR reporting webpage (CAA, 2018). A data request setting out the intended purpose and use of the report data, and guide inclusion and exclusion criteria was subsequently made via the form to the CAA. Since the nature of the request relied on an internal word and category search within the database, a suggested list of descriptor terms was specified.

Intended use specification

To conduct a systematic analysis of occurrences that have been associated with, or cited fatigue affecting flight crew. This research seeks to assess how comprehensively fatigue risk factors may be derived and analysed, and how trends may be determined from the occurrence narratives (and other data fields).

Data request

- UK pilot fatigue-related occurrence reports covering the period 2009-2019
- Covering fixed wing commercial air transport operations
- Output in Excel format with relevant classification fields
- For the same criteria, the annual number of occurrences for all types of event (i.e. not just fatigue-related MORs)

Data inclusion criteria

- Reports applicable to flight crew
- Including:
    - All reported occurrences where the report is deemed to have been caused or related to fatigue
    - Reports citing key phrases and event types associated with fatigue (inclusive of keywords and all other associated word derivatives and phrases such as: tired, tiredness, sleep, falling asleep, sleepy, sleepiness, fatigue, fatigued, exhausted, exhaustion)

Data exclusion criteria

- Events involving other air operations personnel such as maintenance and/or ground handling agents and aerodrome staff
- Non-civil air operations; Rotorcraft operations

Stages 2-3. Request submitted and retrieved from the CAA

The request was submitted using the SRG1605 application form for data release to the CAA with the above specification criteria. The occurrence reports were retrieved via an internal search within the UK MOR database for fatigue and 'all other associated words and sentences such as: fatigue, fatigued, tired, tiredness, exhausted, delays, concentrate, concentration, falling asleep, forgetting, long tiring day, absence, lacking concentration, degraded performance, busy, heavy duty, long hours, poor sleep and workload'. Hence the report data retrieved were stated to be all reported occurrences where the report is deemed to have been caused or related to fatigue.

Data retrieved

- The retrieved full dataset consisted of 2182 fatigue-related MOR reports covering period 2009-2019
- Involving fixed wing aircraft exceeding a maximum take off weight (MTOW) of 5700kg.
- An expansion on word terms specified in data request to include associated words and sentences: 'delays, concentrate, concentration, forgetting, long tiring day, absence, lacking concentration, degraded performance, busy, heavy duty, long hours, poor sleep and workload'.

Stage 4. Coding trial

The initial coding scheme was piloted on a random sample of reports to identify any problems in the ability for it to be applied consistently. Following an initial assessment of a sample of twenty MORs for the inclusion of fatigue risk data metrics outlined in Table 4, the methodology was adjusted to reflect a further binary distinction on data quality and specificity for FRT levels 1 and 2. The basic inclusion criteria was modified to include a further quality distinction between occasions where reports merely mentioned fatigue risk data metrics in passing, or in colloquial / imprecise terms and occasions where full precise timing details or specific details had been provided. Hence MORs were reviewed for whether they mentioned the relevant data category, and if mentioned whether there were sufficient details to enable aggregate analysis of the findings. This distinction is illustrated with an example in Table 4.

Table 4. Coding structure for assessment of coverage in report narratives

FRT Data Metric	Mentioned *Any relevant information referring to data category, even if just in passing*	Details provided *Information relating to precise details in terms of timing or fuller description of details relevant to the data category*
Duty length	i) Reference to length of duty E.g. *'long duty day'* *'extremely long duty'*	i) Indication of length of duty (hours) E.g. *'10 hours on duty'; '15 hour duty day'; ' 'Length of duty including travel time 18hrs 20mins Duty length 16hrs 20 mins.'*

Stage 5. Retrieval of 10% sample of incident reports

A random 10% sample was subsequently generated via a random number generator function in R (R Core Team, 2017) and used to select the reports for the study sample from the full fatigue-related MOR dataset. The initial random sample drawn from the dataset included 21 reports which related to non-flight crew members and one report entry number for which there was missing data. Hence these reports were replaced by a further 22 reports retrieved via the same random number generation method, to produce a final sample of 218 reports.

Stage 6. First Stage Analysis: Assessment of coverage and quality of Fatigue Risk Trajectory levels 1-3

MORs were first assessed for inclusion of fatigue risk data informed by FRT levels 1-3 (Dawson & McCulloch, 2005; Phillips et al., 2017) with additional measures considered of importance for appraising the fatigue safety risk in the aviation setting (see Table 5). At FRT level 1, *duty length* and *duty timing* metrics represent the main categories of data required to assess the adequacy of sleep opportunities afforded by duty schedules with respect to sleep and circadian factors. Information on the time of day the occurrence occurred allows determination of fatigue risk with respect to the two circadian nadirs (Folkard et al., 2006; Horne & Reyner, 1999). Biomathematical fatigue model predictions allow the prior three measures to be used to more accurately capture the duty fatigue risks and likely periods of reduced pre-duty sleep produced by aviation work schedules (CASA, 2014) . This is because they are able to more accurately quantify the sleep-wake parameters from variable shift patterns and circadian desynchrony from travel across time zones, and hence are used widely across the aviation industry.

Table 5. Fatigue risk trajectory framework and corresponding fatigue risk metrics

FRT level	Fatigue risk data metric	Description	Fatigue risk information provided
1. Duty-driven fatigue risks	Duty length Duty timing	The time on duty The time of day the duty occurred	Sleep opportunity / average sleep that the duty schedule provides
	Time of day	Time of day when the reportable occurrence occurred	Determination of fatigue risk at time of occurrence with respect to the circadian cycle
	Biomathematical fatigue model prediction	Fatigue risk estimate of the sleep/wake and circadian dynamics of the prior work schedule	Quantitative prediction of the on-duty fatigue risks and the pre-duty sleep opportunities afforded by the duty schedule
	Duty sectors	Number of flights occurred during the duty period.	An index of increased workload The number of critical phases of flight covered by the duty schedule
2. Individual sleep and circadian factors	Sleep timing Sleep length	The duration and timing of the actual pre-duty sleep achieved by the reporter prior to the reportable occurrence	Determination of actual sleep obtained by the reporter prior to the duty
	Sleep quality	The quality of sleep achieved by the reporter prior to the reportable occurrence	The reporter's qualitative experience of the recovery value of the sleep obtained prior to duty
	Time of Continual Wakefulness	The individual level of continual time awake to the reported occurrence	The duration of the reporter's hours of prior wakefulness provides an indication of their individual acute sleepiness level
	Commute information	Individual commute information	The duration and timing of the reporter's commute
3. Fatigue symptoms	Subjective experience	The fatigue symptoms experienced by the reporter and rating on a validated fatigue scale	The reporter's fatigue related symptoms experienced for the reported occurrence
	Crewmember's experience	The fatigue symptoms experienced by other crewmembers	The fatigue symptoms experienced the other flying crew member/s for the reported occurrence

At FRT level 2, the sleep metrics provide individual level information on the actual sleep obtained and whether this differs from that expected from the duty schedule. The individual wake time and commute information metrics further provide information on the exact hours of wakefulness experienced by the reporter beyond the basic duty schedule for the reported event. At FRT level 3, the individual's experience of fatigue and associated symptoms may be used to approximate the severity of the individual fatigue level. The provision of the reporter's rating on a validated subjective sleepiness scale, such as the Karolinska Sleepiness Scale (KSS, Åkerstedt et al., 2014) provides further interpretative value of the severity of the fatigue state experienced by the reporter as more reliable inferences can made on the likely performance impairment. In aviation, these two latter fatigue risk data categories are furthermore important to establish for the other flight crew members for a safety event to provide a further indication of simultaneous fatigue risk to the human component sub-system.

Stage 7. Second Stage Analysis: Qualitative appraisal of report narratives

The second stage of the review involved a qualitative appraisal of the MOR data. Initial codes were developed based on an examination of the individual report data, and an iterative analysis of the report narratives was undertaken using a combination of inductive and deductive coding techniques. A second researcher reviewed the coding framework against a small sample of reports (n = 10) in order to provide qualitative feedback on its practical use and applicability. Where possible, report themes were organised and developed against the levels of the fatigue risk trajectory framework, and prominent recurring data points (specific words, phrases or overlapping sentiments) were noted across the entire sample to aid the development of thematic categories. For the appraisal of the fatigue related errors and safety occurrences (FRT levels 4-5), the report narratives were first assessed for their existing database categorisation with respect to safety occurrence class and event type. The former classification represents an assessment performed by the CAA to denote the safety severity of the reported occurrence. The latter 'event type' classification was not found to provide sufficient granularity of analysis for identifying specific errors or details of the safety occurrence for use in this study. For example, 81% of the study sample data were classified under the main event category of 'Personnel alertness and fatigue events'. Hence the report narrative data were appraised for the behavioural symptoms, errors and main safety themes emerging from the data alongside any discrete operational events or outcomes that could be discerned from the written text with a priori knowledge of MOR reportable occurrences. These findings were subsequently summarised where possible to provide clarity on the content of the fatigue reports. During the review process, additional notes were made on the data quality and the practical experience of discerning, interpreting, and extracting information for knowledge transfer from the sample dataset.

3.3 Results

MOR Dataset

Figure 7 shows the absolute number of fatigue-related MORs submitted to (and categorised by) the UK Civil Aviation Authority between 2014-2019. In this period, the percentage of fatigue-related MORs increased from 0.01 – 3.22% of all submitted MORs, with an over 600 fold increase in the number of reports submitted. The increase in fatigue-related MORs between 2015-2016 coincides with a regulatory change in flight time limitation rules to EASA regulation (EASA Subpart FTL) brought into effect in February 2016, and some changes to the Occurrence Reporting framework (Regulation (EU) No 376/2014), applicable since 15 November 2015 to aviation professionals, organisations and Member States, on reporting standards for relevant safety information. Hence it is likely that either or both of these factors (increased reporting emphasis or changes to the flight time regulations) contributed to the marked increase fatigue-related reports between 2015 and 2016.

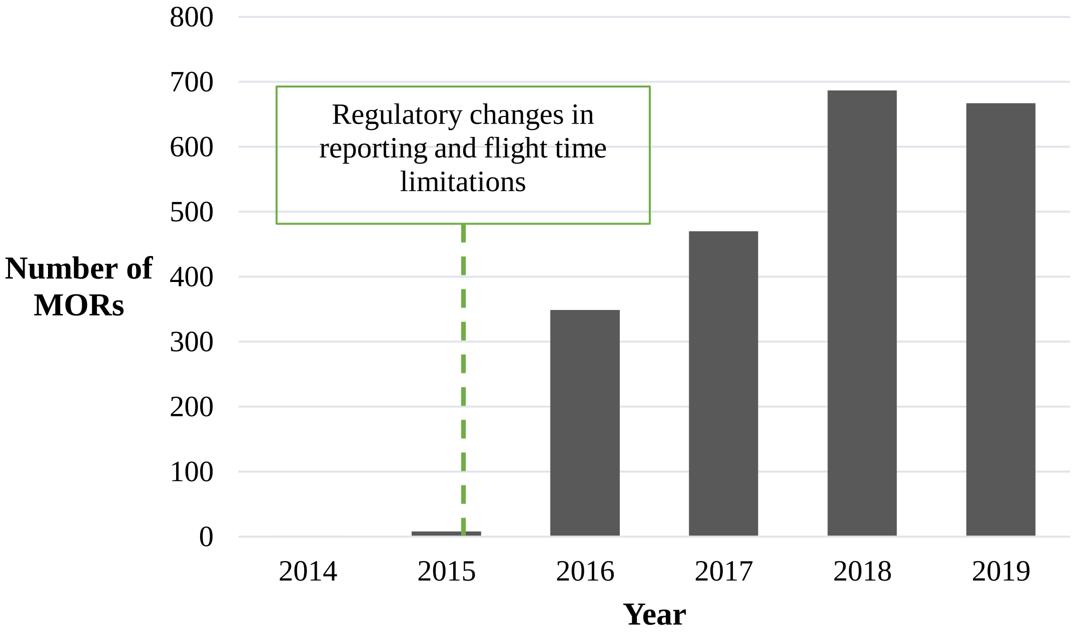


Figure 7. Fatigue-related MORs submitted to the UK Civil Aviation Authority 2014-2019

Study Sample

The following results represent the analysis of 10% random sample (n = 218) of the UK MOR database. As may be seen in Table 6, the final 10% sample did not significantly differ to the full dataset of fatigue-related MORs in terms of proportion of reports by CAA coded flight phase (X^2 (9) = 8.5545, $p = 0.48$.), although the sample did overall significantly differ in terms of year X^2 (5) = 46.021, $p < 0.001$), because there was a higher proportion of reports retrieved in the year of 2015 within the random sample, compared with the proportion expected from the full dataset. However, the proportion of reports expected for other years did not substantially deviate from expected proportions, as such the sample can be seen as broadly representative of the full dataset. No fatigue-related MORs categorised by the CAA were returned for the period 2009 - 2013.

Table 6. Study Sample Characteristics

	MOR Fatigue reports	
Data Category	**Full dataset (n = 2182)**	**Sample (n = 218)**
Year		
2014	1	1
2015	8	6
2016	349	30
2017	470	47
2018	687	57
2019	667	77
Flight Phase		
Standing	512	47
Tow	1	0
Manoeuvring	1	0
Taxi	63	6
Take-off	53	4
En route	644	56
Approach	141	22
Landing	46	3
Unknown	373	45
Not Applicable	348	35

Coverage of duty and individual risk factors in fatigue-related MORs

The first aim of the present study was to assess the coverage of key fatigue risk data categories at the duty level (FRT level 1) and individual level (FRT level 2) in the MOR report data. As may be seen in Figure 8, the overall coverage of FRT metrics in the MORs examined was extremely limited, with less than half of the MORs reviewed containing any mention of level 1 and 2 factors, and an even smaller proportion of MORs providing specific details in terms of timing or features explaining how the relevant data category had an impact on the reporter’s fatigue-related occurrence. With respect to FRT level 1 duty factors, less than 50% of MORs mentioned duty length and timing information, whilst only 7.7% provided actual details of duty start and end times. Hence, the prerequisite timing data needed for post-event bio-mathematical estimations of duty-driven fatigue risks or appraisal of individual sleep-wake factors were not included in the majority of the MOR report data, and just one report in our sample contained both sets of timing information.

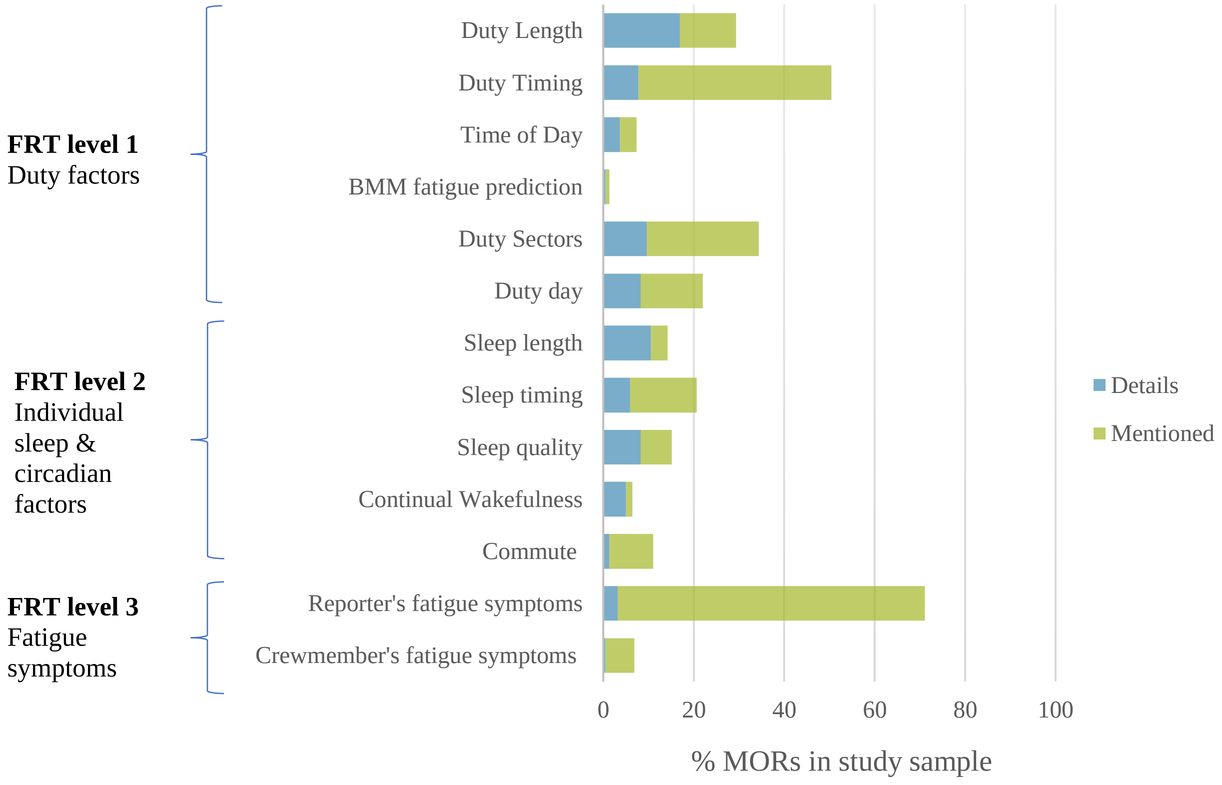


Figure 8. Proportion of MORs containing fatigue risk data at FRT levels 1-3

Within each category, the green portion of the bar indicates the proportion of MORs which mentioned information relating to the category, however briefly or incompletely. Where reports included actual time or specific information relating to the category, this was coded as a report that both mentioned and contained details (displayed in blue). This distinction was used to highlight the coverage of data of sufficient specificity to be able to be extracted for trending purposes.

Two MOR reports included a fatigue model estimate for the incident duty and a brief interpretation of the model output, although in neither case were the exact duty or sleep times detailed, nor the source or exposition of the modelling provided. A considerably greater proportion of reports in the study sample provided some mention of FRT level 1 schedule features (Duty sectors, Length, Time of day) as general context preceding the incident event, with varying degrees of commentary linking these schedule features to their behavioural or operational hassle factors outcomes. For example, 29.8% of MORs mentioned duty length either explicitly as a fatigue factor or as context, with 2.3% reports detailing extremes of work periods extending over 17 hours or extended periods of continual wakefulness as a result of operational factors such as delays or calls from standby, diversions or overtime work. However, for most reports, this information was either not available, was too vague e.g. *'…last sector of a very, very long day, again'*; or contained insufficient details for analysis e.g. *'…series of work had been 6 on 2 off, 6 on and 2 off followed by a further five days work.'*, and overall there was limited consistency between reports in the quantitative details provided and a lack of clarity over their temporal proximity to the reported safety occurrence.

For individual sleep-wake pattern data, only 10.5% of MORs provided specific sleep length data within their report, and just 5% of MORs mentioned both sleep and wake up times experienced by the reporter prior to the incident duty. Where this timing information was provided most reports referred to intermittent or no sleep prior to duty, or between 4.5-6 hours sleep. Where some reference to sleep quality was made (15.1% of MORs), the narratives mostly centred on poor, broken or sub-optimal sleep, with one or more wake ups as a result of circadian de-synchrony, hotel issues (noise, temperature etc.), and poor sleep relating to medical issues. Although only 5% of MORs provided exact details of the reporter's continual hours of wakefulness, where these were provided, the range was between 17.00 -19.15 hours continual wakefulness. Whilst there were greater proportions of MOR reports referencing FRT level 2 data in passing, overall there was limited consistency between report data in terms of the individual sleep and wake data included, and the temporal proximity of these data points to the reported safety occurrence. Taken together, although some reports provided data and specific details on elevated fatigue risks, a systematic quantitative appraisal of trends for either Hazard 1 or 2 levels was not deemed possible, raising questions over what quantitative insights may be drawn over fatigue risks from the MOR broader safety occurrence database.

Qualitative assessment of FRT Levels 1 and 2 data

Against the limited coverage of the specific fatigue risk factor metrics in the MORs analysed, there was nevertheless a considerable degree of thematic overlap between reports over the key work schedule features presented as contributory to the safety event, or as background context to the reporters' fatigue-related report. Notwithstanding the aforementioned data caveats, from a qualitative perspective, of the data nevertheless available there were three common themes related to duty and individual sleep and circadian risk factors which are discussed below. These are summarised in Table 7 (i-iii) alongside a range of quotations from different MOR reports.

i. Insufficient sleep opportunity

The most common FRT 1 factor raised as an issue was the time available for recovery sleep and rest in between duties (Table 7.i). Many report narratives stated that the planned off-duty hours did not afford sufficient time for sleep opportunities prior to or in between duty days. As illustrated in Table 7 i. reports either referred directly to the sleep time available which were less than six hours of sleep e.g. *'...sleep time available of 5 hours and 15mins'* or referenced a particular scheduling rule or practice. For example, 7.7% of MOR reports specifically referred to the fact that the incident duty had been preceded by 'minimum rest', which was seen to be insufficient to obtain sufficient sleep, particularly where there had been repeated consecutive days of minimum rest, or minimum rest was combined with elongated duty days. As a scheduling term within commercial aviation, 'minimum rest' refers to the off-duty period provided being 'at least as long as the preceding duty period, or 12 hours, whichever is greater,' with a reduction to 10 hours off, if the off-duty period occurs away from the pilots' home base (EASA, ORO.FTL. 235). Hence such reports refer to sleep periods where the rostered opportunity available for sleep built into the schedule is by definition at the minimum level, which reporters felt did not provide a sufficient sleep opportunity. Other reports highlighted that last minute amendments to the schedule had substantially reduced the sleep opportunities between duties. Such reports suggested a disparity between the presumed sleep/rest timings of their planned rosters and the actual times available for sleep on the day, once operational hassle factors were taken into account. The problem presented by this is that pilots' expected sleep-wake preparation for duty did not match the sleep-wake preparation the actual duty hours required. Hence, reports detailing insufficient sleep opportunities related to both planned work schedules where the off duty time was felt to be insufficient for recovery sleep and unplanned amendments to schedules where last minute changes to the roster reduced the amount of time available for sleep.

ii. Timing of sleep opportunities

The second theme raised in a number of reports related to issue that the sleep opportunity timing between duties was unsuitable in terms of sleep/wake pressure or circadian drives (examples provided in Table 7: ii). Reports that highlighted a roster issue with circadian desynchrony or ‘body-clock changes’ tended to describe how the out of work gaps between duties were unsuitable in terms of likely sleep drive and timing, given their prior pattern of work hours. In addition, 16% of MOR reports mentioned ‘early to late’ or ‘late to early’ transitions in duty start time between consecutive duty days as a contextual factor within their report. The problem highlighted by reporters is that rapid shifts in duty start times cannot be accompanied by similar shifts in sleep timing from one day to the next, as this is physiologically difficult to achieve.

iii. Cumulative impact of insufficient recovery days off and recurrent sleep loss

The impact of an accumulation of fatigue or sleep dept over a series of work periods was highlighted by a number of reports (Table 7: iii). Often such reports outlined a series of duty days or weeks which included consecutive insufficient sleep opportunities or repeated features of duties outlined in the earlier two themes. For example, 20.6 % of MORs in the study sample referenced consecutive early starts or late duty finishes as context to their fatigue report, approximately half of which (9.6%) also referenced rapid transitions between early and late duty types. Others more colloquially referenced weeks or months of intensive or ‘heavy’ work periods and insufficient sleep and rest recovery periods between blocks of duty which had over time, compounded into greater fatigue levels experienced on duty.

Table 7. Key themes and example quotations from report narratives relating to FRT levels 1 and 2

Key Theme	FRT 1. Sleep opportunity afforded by the roster	FRT 2. Sleep/wake pattern achieved by individual
i. Insufficient Sleep	(MOR 1331) *‘ ... It is impossible to gain sufficient rest immediately prior to commencing this rostered flying duty period.* (MOR 114) *‘... the planned roster left a lot to be desired in terms of a proper break… My alarm was set for 04:30, sleep time available of 5 hours and 15mins.’* - (MOR 548) *‘….have to attempt to achieve two sleep patterns between 0130 and 0500 (27.5 hours), in two separate hotels, with limited opportunity to gain sufficient rest.’* - (MOR 14) *‘ rotation being rostered with an 18-30 rest period and an STD [standard time of departure] of 2220 Local time’*	(MOR 2172) *‘…in bed for 0330z, slept for 1.5 hours before waking up…Went back to bed at 1100z ... Slept only 2 hours…recognised the signs of fatigue.’* (MOR 712) *‘Operated two flights prior to this deep night XXX …. I woke up at around 0800 for both of these flights. After my XXX flight, I actively tried to adjust my sleep pattern by staying awake until 0130L … I did wake later but at 0915 so effectively lost 1h30m sleep. Attempted to recover some sleep between 1600 and 2000 to prepare for the deep night flight but only managed around an hour of broken sleep. Felt tired when I reported for work at 2230.’* (MOR 429) *‘…virtually no sleep before flight’*
ii. Sleep difficulties due to duty timing	(MOR 784) *‘Rest periods were provided at times that were unsuitable. Being required to have two rest periods [and a] transition [from] early to late …[has an] impact on ability to achieve adequate rest;’* (MOR 908) *‘Just a local night to shift one's body clock from lates to earlies’* (MOR 1289) *‘[XXX] should be a stand alone duty and not rostered the next day with anything to secure an adequate rest. Certainly not rostered with a night [XXX] that lands at 7:30am’*	(MOR 1267) *‘…consecutive nights not getting enough sleep due to roster pattern starting on a late with flights starting earlier each day, often with close to minimum rest, led to not enough sleep especially by duty on the XXX. Last 3 duties were extremely early which made adapting from lates to earlier especially hard. Managed to increase amount of sleep for last duty so not as tired as previous duty but still showing signs of tiredness’* (MOR 1282) ‘*Every time I have operated this day … I have either felt very low alertness levels at the end of sector 3 and into sector 4 or completely wiped out the following morning…I was so tired when getting home I was forced to retire to bed at 20:00. This added departure from my normal sleeping pattern meant that I woke up around 2 am and struggled to get further rest.*’
iii. Cumulative fatigue relating to longer term sleep loss or insufficient rest between duties	(MOR 2109) *‘ I have worked a heavy pattern of XXX which is a period of 19 days with 4 days off... These consistent, high density work patterns are not sustainable and should be considered dangerous…’* (MOR 624) *‘…an accumulation of fatigue over the past couple of blocks, with great difficulty adjusting to the rapidly changing early-late-late schedule due to insufficient days off between the blocks (in reserve period), as my previous two fatigue reports show…’* (MOR 1139) *‘Roster pattern such that I feel unable to operate safely due fatigue. 3 days off in 17, switched from early to late flights, plus sim check….’*	(MOR 2069) ‘*…heavy roster over the last 3 weeks and this week in particular with its very long days with big delays and late finish time …I was going to bed VERY late after these duties…. my body clock seemed to be a bit confused and didn't end up going to sleep until 0200 and was needing to get up at 8 for another 11.30 report… due to the nature of the shifts in the last block [and] the length of the shifts over the 3 blocks and the fact of only having 2 days off between each all lead to fatigue’* (MOR 1411) *‘4 early starts in a row with Training XXX mixed in. 14 sectors in 4 days on earlies including delays. This was followed immediately by 2 late sectors. This resulted in mixed up sleeping pattern and operating whilst tired... Feeling extremely tired on days off. No chance to properly recover before 4 sector days after only 2 days off’*

Fatigue Risk Trajectory Level 3 - Behavioural fatigue symptoms

Overall, 71.1% of MORs referenced overt behaviours or subjective symptoms relating to the reporter's experience of fatigue, but only 3.2% of MORs provided the severity of sleepiness level as measured by a standardised scientific scale. As expected, the most commonly reported symptoms included in the report narratives related to 'sleepiness' ('struggling to stay awake'; reduced alertness; tiredness; sleep-driven fatigue) which were mentioned in half of the reports (50.5%). However, in 15.6% of these reports, the description provided was third hand, and on most occasions, simply stated that the crew member felt 'reduced alertness' or a 'reduction in alertness' with no further details. Hence it is not entirely clear whether these latter descriptors reflect the reported experience of the reporter or were proxy terms for categorisation purposes. 14.2% of MORs referenced attentional difficulties (including references to difficulty maintain concentration, feeling a lack of focus or distracted during monitoring). 10.1% of MORs referenced feelings of exhaustion (including references to feeling mentally or physically drained, low energy, feeling 'wiped out').

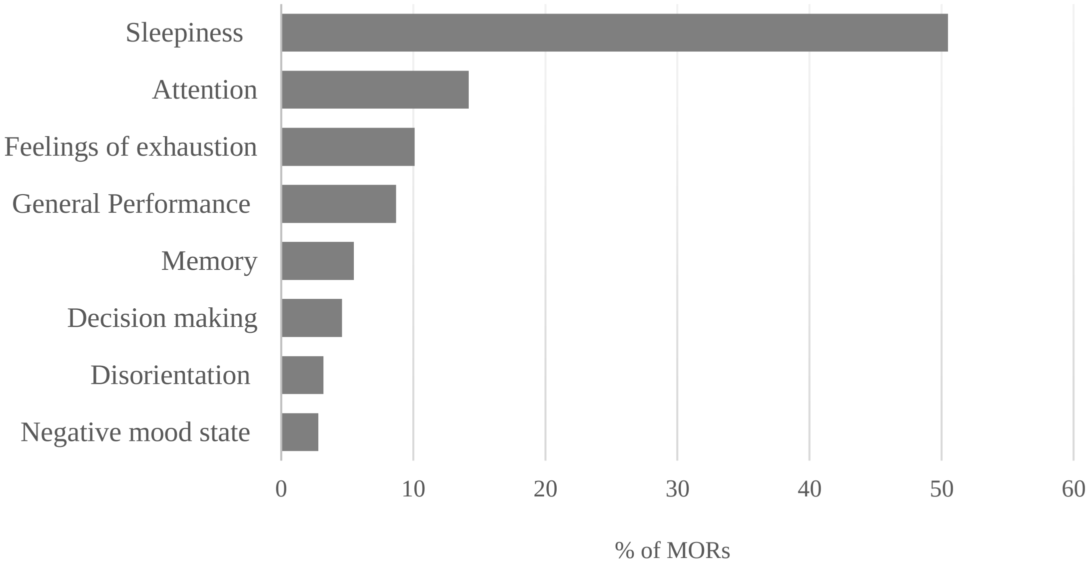


Figure 9. Fatigue symptoms mentioned in MORs

Other fatigue symptoms mentioned in less than 10% of the MORs included aspects of general performance capacity (including references to sub optimal or deteriorating general performance, being at or beyond one's 'capacity' or 'limits'; 'unable to cope'); memory decrements (references to forgetting things, or forgetting to do tasks), reduced decision making ability (references to poorer problem solving and judgements); disorientation (references to feeling dizzy, ‘spaced out’) and negative mood (references to low mood, agitation and frustration). Overall there was a limited consistency between reports as to the description of symptoms with respect to severity levels, their duration and precisely when they occurred with respect to the reported event or concern.

Fatigue risk trajectory Level 4 - Fatigue-related occurrences

Mandatory reportable safety occurrences by definition include not only pre-defined reportable occurrences compromising the safety of the aircraft or operation, but also breaches of technical safety margins or situations in which a safety judgement has been made by the reporter. For this study, appraisal of the FRT level 4 (fatigue-related errors or occurrences) the report narratives were assessed for themes in the types of safety events, errors or concerns raised in the narrative data. Table 8 sets out a thematic summary of the main occurrences and indicative examples derived from the content of the report data.

Table 8. Safety events and concerns in fatigue-related MORs

Safety occurrence	**% of MOR sample**	**Summarised event or concern**	**Example quotations**
i. Events			
In-flight sleepiness /sleep incapacitation	36.2	- Sleepiness at some point during flight - Sleepiness throughout flight - Involuntary sleep event during flight - Sleepiness / sleep event during critical phase of flight	**(MOR 254)** *‘Whilst the captain was asleep, I found myself unable to stay awake… awoken by a radio call.’* **(MOR 498)** *‘Very hard to stay awake in descent - consider fatigue to be of dangerous level….’*
Unfit to fly due to fatigue	23.9	- Fatigued prior to duty - Unfit to continue duty - Commander’s discretion not to extend duty	**(MOR 1471)** *‘...I wasn't fit to operate due to fatigue…it took a lot of courage to make the phone call...’* **(MOR 576)** *‘...I felt that discretion was unjustified. [Its use] should be exceptional and is only for unforeseen circumstances’*
Go-around	3.7	- Go-around required as safest course of action	**(MOR 1630)** *‘Go around flown due to a loss of SA [situational awareness] and not being in the landing configuration’*
ii. Errors			
Call errors	10.6	- Missed/ forgotten calls from Air Traffic Control (ATC) - Read back errors in radio transmissions	**(MOR 1739)** ‘*During last sector signs of fatigue evident. Read back errors in RT [radio transmission] during periods of high workload.*’

		- Missed or delayed standard calls - Missed or failure to complete checklist items	**(MOR 1402)** *'…feeling the signs of FTG [fatigue]…Missing radio calls and some standard calls'*
Programming / selection errors	11.0	- Misselections of altitude - Incorrect setting of flight management systems - Incorrect flap selections on approach	**(MOR 21)** *'…errors setting flight guidance panel and navigation box due [to] fatigue.'* **(MOR 62)** *'…GPWS call "too low flaps" was heard…flap selection switch was at 15 and not at 35 as should be.'*
Control of aircraft	5.5	- Airspeed approach profile errors or deviations - Trajectory (Roll/ Pitch) errors - Landing gear not configured	**(MOR 662)** '*We forgot that we were on manual thrust and the A/C slowed down to almost Vref30+20'* **(MOR 494)** *'…errors during the approach, late selection of gear…'*
iii. Concerns			
Safety concern with duty/ roster	26.6	- Legality of the roster (length or timings or both) - Planning of the duty or roster - Lack of sufficient rest period - Pressure to fly during duty	**(MOR1 514)** *'Whilst these duties may be legal, they are not sensible.*' **(MOR 422)** '*By the time I came off duty…it would have been 19h15m since the standby started... it is pushing the margins of safety…*'
Safety concern over fatigue state	12.8	- Fatigue identified as a threat to safety during the flight - Crew's perception of an inability to respond in an emergency / if conditions were different	**(MOR 1925)** *'Fatigue played a role in the reduction of overall flight safety'* **(MOR 25)** *'…I was exhausted. Just hoped no abnormal situation occurred as I struggled to focus.'*

As may seen in Table 8, the fatigue-related MORs in the study sample included not only reports citing specific proximal on-duty errors or events (Table 8: i - ii) but also a high proportion of reports citing safety concerns over the organisational management of schedule-driven fatigue risks, and the reporter's perception of the reduction of safety margins due to fatigue (Table 8: iii).

i. Events

In-flight sleepiness represented the most commonly reported event (36.2% of MORs in study sample). For the vast majority of these reports, there was no mention of any specific operational consequence of the sleepiness event on the safety of flight. Rather the purpose of the majority of these reports appeared to be to highlight the fact that at a systems level, the safety buffer of having an alert pilot had been compromised on this particular occasion. Whilst only 4 reports (1.8% of MORs) explicitly referenced discrete sleep incapacitation events (involuntarily falling asleep whilst on the flight deck), 2 of which indicating simultaneous incapacitation of both pilots, a much greater proportion of the study sample contained indirect or euphemistic references to involuntary sleep propensity during flight, e.g. *'…could have easily went into a deep sleep if I'd shut my eyes'* (MOR 1649). Hence the true incidence of sleep incapacitations even within these report data is likely to be greater than the number explicitly stated. The state of alertness of the other crew member was only mentioned in 6.9% of all MORs in the sample, but where mentioned, the reports all indicated that both flying pilots were simultaneously affected by high levels of sleepiness, e.g. *'…at various times during the flight I felt … effort required to stay awake (KSS 8)… the Captain …too raised concerns that he also felt sleepy'* (MOR 494). Whilst some reports citing in-flight sleepiness referenced a time point or phase of flight to help denote the safety significance of the event, e.g. '*Very hard to stay awake in descent - consider fatigue to be of dangerous level….*' (MOR 498), for most reports, details on the timing and duration of sleepiness symptoms, phase of flight and sleepiness status of other flight crew were missing.

MORs referencing the reporter's decision not to fly due to safety concerns with fatigue represented the second most commonly reported event in 23.9% of MORs in the sample. Such events relate to a pre-emptive action taken by the reporter as a consequence of their perception that their own fatigue levels might compromise safe performance, or that, in their view, the schedule-driven fatigue risks of the upcoming flight were not reduced to an acceptable level of safety. Since the event relates to a pre-flight decision, the report information tended to focus on the contextual factors leading to this decision. As highlighted in the previous section, precise duty and sleep-wake history information was not present in the majority of reports. However, in over half of the reports citing this safety decision, the reporter mentioned operational factors extending the time or changing the time period required to be awake on-duty (diversions, delays, late call outs from standby) as well as sudden changes to the reporter's roster as context to their decision to not accept the fatigue risks of the upcoming flight.

ii. Errors

11% of MORs detailed errors involving the programming and correct selection of settings in the flight management system, and in the configuration of the aircraft during the take off or approach/ landing critical phases of flight. Over half of these were flap selection errors, where the flap selection was incorrect for the airspeed limit or was performed out of sequence during take off or landing. This type of error if uncorrected leads to altitude and speed deviations in the flight profile of the aircraft. In most reported cases of this type of error, the issue was quickly noticed and rectified by the flight crew, or discovered via a configuration test soon after (as part of a standard operating procedure). Other selection errors predominantly related to mistakes in the setting of the flight management system (incorrect inputs or deletion of inputs) during the approach phase that affected concurrent task management activities over the control of the aircraft trajectory. A similar proportion of MORs (10%) cited communication errors in air traffic control calls and standard operating procedure calls on the flight deck. Such errors predominantly were of omission, relating to missed or incomplete calls and requests and subsequent delays in response, and some mistakes in the repetition of calls.

iii. Safety concerns

There were two main themes of safety concerns raised in the MOR sample**.** In 26.6% of the MORs in the study sample, reporters raised safety concerns that were characterised by the perception that the schedule-related fatigue risks had not been reduced or controlled to an acceptable level. In particular, the concerns highlighted were those over the organisational planning of the work pattern in terms of its length or rest breaks, or permissiveness of the rules that could have allowed the work schedule to be constructed from a safety perspective. In one quarter of MORs where this concern was raised, the pilot took the decision not to continue their flying duties (referenced earlier). However, some MORs also suggested that the reporters felt pressure to fly despite their perception of the fatigue risks. The second type of safety concern related to the reporter's perception that if the conditions of the flight had been different or an emergency situation occurred, they would not have been able to respond appropriately due to their high level of fatigue (e.g. Table 8. MOR 25). Few reports explicitly specified the reasons why despite elevated fatigue risks, a serious incident or accident, in the reporter's view, had not occurred. Those that did predominantly cited timely intervention from or cooperation with the other flight crew member. However other reasons included fair weather conditions (so that there would be reduced consequences of any fatigue-related performance decline), strict adherence to standard operating procedures (aiding management of fatigue-related performance), alerts from air traffic control (e.g. Table 8. MOR 254) or automated warning signals (providing some degree of feedback to the pilot that certain procedures or actions had not been undertaken).

Data quality

In general within the sample of MORs assessed in this study;

1. Reports lacked clarity over the reasons why the report was filed
2. Reports lacked key details on fatigue risk metrics on the reporter's fatigue level (as indicated by a scientifically validated scale) or prevailing sleep-wake history context.
3. Reports lacked consistency in content in terms of information included or excluded, making extraction of key details difficult, vulnerable to misinterpretation or not possible.
    a. Descriptions of tiredness/fatigue symptoms or duty patterns were often generic, euphemistic or non-specific, making it difficult to see how information could be used for 'trending purposes'
    b. Reports within the MOR dataset contained a mixture of 1st hand account (57%), 3rd hand analysis/opinion (22%) or a mixture of both (21%), without explanation over the transposition of raw data, leading to a lack of consistency and clarity across the MOR dataset.

3.4 Discussion of Study 1 Results

Within commercial aviation, the purpose of close call reporting is to learn about *how* and *why* breaches to safety thresholds have occurred in order to prevent future accidents. The Civil Aviation Authority via its mandatory occurrence reporting (MOR) system more specifically monitors trends in the frequency and severity of safety occurrences and associated underlying risk factors that have the potential to threaten flight safety (CAA, 2018). However, the presumed value of such safety reporting systems more generally is not just in understanding the particular constellation of conditions and circumstances surrounding a single safety event, i.e. the *Safety I* approach discussed earlier. Rather, in theory, such systems may also provide a broader aggregated database that over time, provides insights on trends in relationships between underlying risk factors, technological systems, organisational processes and human behaviours associated with safety occurrences (Underwood & Waterson, 2014).

Flight crew fatigue is a widely recognised human factor risk in commercial aviation, and has been identified as a causal or contributory factor in near-miss incidents as well as fatal accidents (Marcus & Rosekind, 2017; National Transportation Safety Board, 2000, 2001, 2010). Survey data also suggest that high proportions of commercial pilots (60-90%) have at one point in time or more regularly, experienced fatigue-related decline in their piloting performance (Aljurf et al., 2018; European Cockpit Association, 2012; Gregory et al., 2010; Reis et al., 2013). Whilst prior research points to an overall positive correlation between hours on duty and probability of an in-flight incident (Goode, 2003; O'Hagan et al., 2016), to date, surprisingly little appears to be known about the relationships between increased fatigue risks, single or multi-crew pilot performance impairment and the emergent impact of such fatigue risks to safe commercial flight operations. Although pilot fatigue reports constitute the crucial substrate data input into an individual organisation's FRMS, prior survey data has suggested not only widespread underreporting of fatigue events by European flight crew via formal company channels, but also negative safety culture perceptions around the organisational management of the fatigue hazard (ECA, 2012; Reader et al., 2016). Within the context of this PhD, the study of the industry's central safety occurrence database hence represented an essential starting point for understanding how fatigue risks are captured in real-world safety occurrence data, and whether there were emergent themes that characterised fatigue-related occurrences that could be systematically appraised. Hence, in order to investigate how fatigue risks may erode critical safety margins during commercial aviation operations, Study 1 sought to

1) Assess the quality and coverage of fatigue risk data currently captured by the MOR system
2) Investigate associations between fatigue risk parameters and increased risks to pilot performance and emergent system safety and
3) Evaluate any emergent insights, relationships, common themes or characteristics that may be gained from an examination of fatigue-related MORs. In light of these assessments, quality improvement opportunities for data gathering, fatigue risk monitoring and future reporting requirements were furthermore considered as part of the study.

Overview

Study 1 of fatigue-related mandatory occurrence reports suggested that overall there was poor and inconsistent coverage of fatigue risk data included in MORs, where fatigue was cited as a predominant or contributing factor. In general, the report data in the study sample lacked consistency in terms of the basic fatigue information included or excluded. Where included, often fatigue risks were described colloquially or briefly in passing, making extraction of objective details difficult, vulnerable to misinterpretation or not possible. Against the Fatigue risk trajectory framework (Dawson & McCulloch, 2005), MOR narratives did not include sufficient data to allow for fatigue risks at any trajectory layer surrounding the reported safety occurrence to be quantitatively investigated and compared between reports.

Quality and coverage of FRT 1-2 fatigue risk data within MORs

Most reports mentioning work pattern or individual sleep-wake history information (FRT levels 1-2) did not include specific timing information (sleep-wake and work schedule times and the reporter's state of acclimitization to the local time zone) prior to the reported event or concern. At the individual report level, where such quantitative information was included, there were data examples of fatigue risks that would be considered as elevated in a safety-critical shift work context, such as the 10.5% of MORs which reported sleep lengths that ranged between 0-6 hours sleep prior to flight and the 5% of MORs which provided exact details of the reporter's continual hours of wakefulness that ranged between 17.00 -19.15 hours continual wakefulness. However, the combination of precise duty information, sleep history and point in time of fatigue-related occurrence were rarely provided, considerably limiting the quantitative appraisal of fatigue-risks for individual reports. Consequently at the aggregate level, lack of precise data and consistency between information included in reports meant that the prerequisite data needed for larger scale bio-mathematical estimations of duty-driven fatigue risks was not possible. As such, across the study sample there was insufficient data to allow systematic quantitative comparison between reports over key preceding work schedule and sleep history (FRT 1- 2) fatigue risk parameters relating to safety occurrences.

This particular finding of insufficient specificity of duty and sleep times contained within aviation safety occurrence reports is consistent with the findings of Lyman and Orlady (1981), who three decades earlier, examined the Aviation Safety Reporting System (ASRS; covering voluntary safety reporting in the United States) for fatigue-related aviation incidents. Whilst duty period, sleep and rest factors were identified as three of the most frequently cited factors leading to 'aircrew fatigue' in ASRS reports, the authors concluded that '*…the information presently within the ASRS database does not permit an analysis in depth of the effect of such factors…'* (Lyman & Orlady; 1981, p20) and within their analysis noted the particular problem for aviation operations in the recording of accurate time information related to person's "body time" verses the local time for which the incident might take place (Lyman & Orlady; 1981, p10). Advances in bio-mathematical fatigue modelling during the 1980s and 90s have subsequently allowed applied sleep science researchers to better approximate the relative circadian disruption that may result from transmeridian travel, large shifts in duty start times or prior sleep-wake histories where these data are available (Dawson et al., 2011; Rodrigues et al., 2020). Hence, even where flight times are provided in reports, the context of the pilot's prior shift pattern and home base is necessary to establish their predicted fatigue risk as well as the local time zone. In a later study of the ASRS system assessing fatigue in air traffic controllers, similar conclusions were drawn, and the authors underscored the research need to assess duty schedules and shift-related factors with more specific fatigue-risk information (della Rocco, 1999). One notable difference between the ASRS and MOR systems is that the latter is a mandatory reporting system, where the quality of data might be expected to be greater owing to the significance of submitted reports within the safety regulatory framework. Whilst there do not appear to be similar published peer reviewed studies of the MOR system for fatigue, a recent technical report for the CAA seeking the analysis of accidents and incidents involving flight crew fatigue (Roelen & van Dijk, 2019), notably used data that was 'mostly from the NTSB', which is the main transport safety standards body in the United States. One reason for this captured in a footnote cited the superior 'quality' of the NTSB incident and accident investigations (Chapter 3, p38), which perhaps suggests that other researchers have encountered similar difficulties in the retrieving of sufficient quantitative fatigue risk information from the MOR database for their research purposes.

This gap in data coverage suggests that the current mandatory data requirements for mandatory occurrence reporting do not elicit sufficient common fatigue risk details for thorough individual event appraisal or broader trending analysis across the wider MOR dataset using quantitative data. It is likely that a lack of industry consensus over the key required metrics for fatigue risk reporting within MORs is a major contributing factor underpinning the low coverage of contextual fatigue risk data. As one indication of this, the narrative accounts in the MORs did not adhere to a common structure or yield a consistent set of fatigue risk terms or information. In many cases, reporters colloquially referred to work patterns that reflected specific individual company scheduling practices, without precise timing data or specification of the particular regulatory rules that had permitted the work schedule. Within such narratives, schedule-driven fatigue factors were frequently mentioned. For example 16% of MOR reports cited 'early to late' or 'late to early' transitions in duty start time between consecutive duty days as contributing to the reporter's fatigue status. However, further description of precise details of these 'early' and 'late' times or the degree of circadian disruption was not included. A small number of MORs also mentioned that further fatigue information was captured by a separate company report or assessment, the details of which were not included within the MOR report. Against these broader findings, it is possible that reporters were not aware of how the schedule- and sleep history fatigue data provided within MOR reports are collated and appraised by their individual company and subsequently transferred to the Civil Aviation Authority database.

However, with respect to qualitative analysis of FRT levels 1-2 within the data, report narratives nevertheless suggested considerable overlap in broader distal themes associated with duty and sleep factors. These included insufficient sleep opportunity prior to flight, sleep difficulties due to duty timing, and insufficient recovery time accumulating between duties. In particular, the perception of inadequate pre-duty sleep opportunities (as a consequence of their length and/or timing) were recurrent themes linked with the reporter's reduced or degraded sleep and on-duty fatigue. In most cases, the report narratives suggested that such factors were in-built and inherent features of the work pattern, rather than having occurred due to home-life or incidental factors. However, unplanned amendments to schedules were also highlighted as a driving factor for insufficient sleep and disruption to the reporter's sleep pattern. A third major theme was that of a cumulative context of consecutive work patterns that had interfered with the reporter's sleep patterns over extended timescales, that substantially impacted the reporter's ability to recover effectively in-between duties, even on days off. Thus despite lacking the precise timing details necessary for a quantitative assessment of these data, the sample provided thematic convergence on these distal fatigue risk factors and furthermore underscored the value of future data collection of precise roster details and sleep-wake histories from reporters.

Reported fatigue experience (FRT level 3)

As conceptualised by Dawson and McCulloch (2005), at FRT levels 3-4 the risk information relates to the fatigue-related symptoms and effects on performance actually experienced by the individual whilst on duty. Hence such data represent more proximal events on the fatigue risk trajectory and an increasing threat to overall system safety (since more distal defences have not prevented a fatigued pilot from operating). In terms of industry-specific safety learning, these data hence are important to capture within MORs to better understand not only the real world relationships between fatigue and reduced pilot performance, but also fatigue-related pilot performance and safety risks to the overall aircraft operation.

At the aggregate level, the main cognitive symptoms reported in MORs related to alertness (e.g. ‘struggling to stay awake’, ‘tiredness’), and attention (e.g. ‘difficult to concentrate’, ‘unable to focus’ or ‘distracted’). These findings of the characteristics of fatigue-related neurobehavioural symptoms are in line with the broader sleep scientific literature on the hallmark symptoms of sleep-driven fatigue (Durmer & Dinges, 2005; Goel et al., 2009; Lim & Dinges, 2010), and also reflect prior survey research conducted with commercial pilots highlighting alertness, attention and lack of concentration as the most frequently experienced fatigue-related symptoms in-flight (Bourgeois-Bougrine et al., 2003). MOR narratives that cited general piloting performance decline (9%) may also be referencing a general difficulty in managing flying tasks which require directed and sustained attention (Bourgeois-Bougrine et al., 2003). However, in general, such symptoms were often mentioned in passing, with no precise reference to symptom severity, and limited detail over the time period or phase of flight for which symptoms were experienced, leaving the resultant impact of the pilots’ fatigued state on safe flight unclear. Only 3.2% of MORs referenced a scientifically validated scale rating of severity for the alertness / sleepiness dimension of the reporter’s fatigue experience, which meant that there was no consistent measurement unit denoting the severity of symptoms that could be compared across reports for the reporter’s sleepiness level either at the time of the reported occurrence or preceding the occurrence. In addition only a very small proportion (6.9%) of the reports included reference to the concomitant fatigue status of the other flight crew member. Such information is clearly of crucial safety relevance for understanding whether at any point during flight there was ‘dual incapacitation’ due to fatigue across both flight crew, threatening the entire safety-critical pilot sub-system of the operation. A further issue in reviewing these data for fatigue-related symptoms and subjective experience was that for 15.6% of the MOR sample, the reports were written in third hand. For these reports, the description of the pilot’s fatigue state on most occasions simply stated that the pilot felt ‘reduced alertness’ or a ‘reduction in alertness’ with no further details. Hence for those 15.6% of reports, it was not clear from the data whether descriptions of fatigue-related symptoms truly reflected the experience of the reporter or were in fact proxy terms for categorisation purposes at some point within the MOR submission or review process. Subsequent personal communication with staff contacts with experience of the MOR system at the CAA suggested that the latter explanation was more likely.

Together these findings concerning fatigue-related symptom data (FRT 3 risk coverage) highlight a major improvement opportunity for future MOR data collection, and suggest the need for additional mandated and standardised reporting metrics over sleepiness measurement, pilot sub-system and whole-system safety risks to inform the principle regulatory database. In terms of the coverage of subjective experience of fatigue**,** it should be noted that there is no perfect scientific measurement tool for summarising the nature of the fatigue experienced by the reporter (Phillips, 2015). Moreover, there is undoubtedly continued value in collecting the idiosyncratic fatigue related experiences of individuals in their own descriptor terms that can be evaluated on an individual case basis, particularly where new phenomena might be reported. For example some studies with military pilots have uncovered interesting dissociations between fatigue-related cognitive decline and performance on flight tasks, where isolated aspects of piloting performance may be preserved (e.g. accurate instrument scanning) even when there is concurrent evidence of significant visual perceptual impairment, visual neglect and reduced flying precision (Russo et al., 2004; 2005; Previc et al., 2009) However, since the review of fatigue-related symptoms in commercial pilots most frequently referenced alertness decline, it is argued that the inclusion of the KSS to denote severity of sleepiness experienced would provide an improved degree of scientific understanding of the level of neurobehavioural decline experienced and comparison between report narratives. Such inclusion criteria should not prevent additional reporting of symptoms for MORs where operational errors are captured, but rather allow greater understanding of the relationships between severity of experienced sleepiness levels, performance decline and operational errors.

Safety content of fatigue-related MORs (FRT level 4).

The analysis of the safety content of fatigue-related MORs revealed an interesting mixture of data insights that challenged traditional incident classification and interpretation. Whilst perhaps unsurprisingly, the primary ‘event’ cited in fatigue-related MORs was in-flight sleepiness or sleep-driven incapacitation (36.2% of MORs within the study sample), a strikingly high proportion of the entire study sample cited safety concerns or events that were not associated with specific detectable errors, failures or proximal operational consequences. Instead, many reports provided descriptions of severe fatigue which, from the reporter’s perspective, appeared to represent a critical erosion of the assumed, yet ‘invisible’ safety margin of alertness in one or both flight crew, despite there having been no visible consequence of this reduction in safety during flight. For example, some report narratives suggested that the pilot felt there would have been serious safety consequences of their fatigued state, if an emergency or non-normal situation had arisen, an assessment made explicitly in 12.8% of MORs. It appears that the crucial safety principle and concern highlighted by such reports is that the safety of flight on the reported occasion critically depended on other safety-critical sub-systems not failing, and non-normal conditions not arising. Put another way, from a 'swiss cheese' model perspective, these reports suggested that the safety margin provided by having two or more alert flight crew had been heavily eroded and as such, was not an effective safety buffer. In 26.6% of MORs, safety concerns without explicit operational consequences related to the organisational processes that had permitted or created elevated schedule-driven fatigue risks. The prevailing safety 'warning' signal captured by such reports was that distal organisational factors associated with controlling schedule-driven fatigue risks (including the design of the work schedule itself) had not been sufficient to protect the individual reporter from the proximate situation of elevated fatigue risk resulting from their roster pattern. Hence the nature of these concerns appeared to be that at a systems level, the assumption that schedule-driven fatigue risks were adequately controlled had been violated. Since such distally created risks are not unique to individual pilots or flights and could readily reoccur, it was these factors, over and above any proximate outcomes, that represented ongoing major systematic threats to flight safety. The relatively large proportion of fatigue-related MORs citing organisational concerns over work pattern construction, suggests further data is required within safety reporting to adequately capture the circumstances where the protective safety buffer of flight time regulations has been compromised.

Few reports explicitly specified the reasons why despite elevated fatigue risks, a serious incident or accident, in the reporter's view, had not occurred. Those that did predominantly cited timely intervention from or cooperation with the other flight crew member. However other reasons included fair weather conditions (so that there would be reduced consequences of any fatigue-related performance decline), strict adherence to standard operating procedures (aiding management of fatigue-related performance), alerts from air traffic control or automated warning signals (providing some degree of feedback to the pilot that certain procedures or actions had not been undertaken). One clear possibility is that the reported event of 'unfit to fly due to fatigue', denoting a pilot's decision to not undertake the upcoming flight, representing 23.9% of the MORs in the study sample, is a principle defence layer underpinning why in many occasions, a more serious incident did not occur. Thus despite the high proportion of fatigue-related MORs without proximal operational failures contained within the study sample, it was often not evident *why* the fatigue-related risks did not translate into more substantial safety incursions. Data capture around these protective system features or adaptive individual behaviours may often not be recorded in adverse occurrence reporting databases and require additional research and interviews with frontline staff (Cooke & Rohleder, 2006; Dawson et al., 2017). However, such information clearly is of key safety relevance since it would highlight the existing system defences that might have provided whole- or sub-system protection where a pilot was operating under increased levels of fatigue.

There were smaller proportions of MORs for which performance errors or operational events did occur and where these, if uncorrected, had the potential to affect safe flight. Within the study sample, these predominantly related to selection, procedural or communication errors. For example, 11% of MORs contained errors involving the selection of settings in the flight management system, and in the configuration of the aircraft during the take off or approach/ landing critical phases of flight. Selection errors mainly related to mistakes in the programming of the flight management system (incorrect inputs or deletion of inputs) such as misselections of altitude level on approach, or flap selection errors on approach (which involve a manual pull action on a lever akin to a gear stick). A similar proportion of MORs (10%) cited communication errors in air traffic control calls and standard operating procedure calls on the flight deck. Such errors predominantly were of omission, relating to missed or incompleted calls or requests and subsequent delays in response, and some mistakes in the repetition of calls. These findings on the main types of observable errors contained in fatigue-related MORs are broadly consistent with pilot surveys on in-flight fatigue related errors. For example, as part of their survey work the European Cockpit Association found that ‘wrong setting of switches, using wrong data for performance calculations, failure to follow procedures in safety checklists, and miscommunications or missing air traffic control instructions’ were key examples of fatigue-related mistakes (ECA, 2012). These data, whilst proportionally rarer within the sample, nevertheless provide some suggestive insights on the piloting tasks (configuration of aircraft on approach and communication) and associated cognitive lapses (errors of omission and commission) that are potentially vulnerable to fatigue-related performance decline in normal commercial aviation operations.

3.5 Strengths and Limitations

The MOR database represents the primary repository of real world safety occurrence data in UK commercial aviation. The reporting system itself exists as a key pathway for front line operators to report on safety breaches and for the regulator to identify emerging risks that may threaten flight safety. As such, MORs bear a unique and important safety function as part of the safety regulations governing the UK aviation industry. Against this context, research on the links between human factors risks and safety outcome phenomena should begin by examining the quality and coverage of data captured by the system. Whilst limitations of fatigue risk information collected by safety report systems have been previously identified (Lyman & Orlady, 1981; Rocco, 1999) this study benefitted from using the Fatigue Risk Trajectory as a framework for systematically examining where such gaps in risk information existed at the outset of the analysis. This research approach uncovered an important finding that valuable distal and proximal fatigue-risk information relevant to safety occurrences is often not captured, is difficult to quantitively evaluate at the individual report or aggregate level, or is lost in translation (between primary accounts and third hand appraisals of events). Use of the FRT framework in the review of coverage of fatigue risk information helps to explain why links between pilot fatigue, performance errors and safety occurrences have been historically difficult to appraise quantitatively at the national level. Hence a key benefit of the study methodology was that it provided a systematic understanding of the key weaknesses in fatigue-risk information collected at each FRT risk layer, within a framework that can also be used for future quality improvement in the collection of data.

A second strength of the methodology was that it involved the review of a sample of 218 fatigue-related reports, representing 10% of the total fatigue-related MORs submitted to the CAA over the period 2014 - 2019. In the recent aforementioned technical review of CAA MORs (Roelen & van Dijk, 2019), the methodology adopted excluded reports '…where fatigue was cited but did not lead to any further events' (p 42) and hence only events deemed to have high to medium safety severity (in terms to damage to aircraft or serious increase to flight crew performance burden) were included in their analysis, which yielded a sample of thirty-five MORs. For many reasons, it might be tempting to discount or downgrade the importance of MORs raising safety concerns for occurrences where proximate safety outcomes have not occurred. Yet for some time it has been acknowledged that the exclusive study of observable failures or errors for safety learning is inadequate, particularly in ultra-safe industries which are characterised by complex sociotechnical systems, dynamic control processes and operator involvement which together support and maintain safe system performance (Grant et al., 2018). Hence, it is suggested that both the experience of elevated fatigue levels during flight and the perception of operating pilots that safety margins have been eroded denote important safety threshold data, that should be the subject of research. Indeed, this 'early warning' philosophy is also recognised in the MOR regulation itself which requires the report of any severe in-flight fatigue to be reported, regardless of operational consequence. Such data represent warning signals that can currently only be informed on by the frontline proximal operator, but which nevertheless relate to assumptions over the likely performance failure of the human sub-system. Furthermore, the present study findings suggested that MORs without explicit failures or operational consequences represented a high proportion of all fatigue-related safety occurrences, and hence may contain more representative safety signals relating to fatigue risks within commercial aviation and precursor circumstances than the rarer occasions that have translated into a more serious outcome.

Particular methodological limitations in this study relate to the data retrieval process which involved an internal search within the UK MOR database for fatigue and common synonyms ('exhausted'; 'tired' etc.) or causes ('long hours'; 'poor sleep' etc.). Retrieval of fatigue-related MORs was hence reliant on the basic coding conducted by the CAA and the dataset retrieved may therefore have missed fatigue-related occurrences where the fatigue-risk information was hidden within the report or not cited or emphasized by the reporter themselves. For example, the content of some reports may have contained information pertaining to elevated fatigue risks from a sleep scientific perspective, but the specific word search terms did not correspond with the content of the report, or the MOR itself was not evaluated to be 'fatigue-related' if this was not volunteered by the reporter. A second weakness of the report data was that a proportion of MORs were written in third hand, suggesting that at least for 15% there was a separate report upon which the MOR data captured by the system is based. Original reports were requested but were not available for this study. This means that there may have been some degree of translation error, misunderstanding or bias between the original report and the third hand account – a weakness that appeared to particularly affect the review of fatigue-related symptoms (FRT level 3 risk information).

Conclusion

As with any post-hoc study of safety reports, insights from the present study should serve as the basis for identifying emergent trends in safety issues and also provide new focus for future research. A major, albeit unexpected finding for this study was that fatigue-related safety data is currently poorly and inconsistently captured in the mandatory safety reports that inform the regulatory safety occurrence database. Hence, from the perspective of fatigue-risk surveillance activities across the aviation industry, the ability to assess potential trends in fatigue-risks that translate into safety breaches across UK commercial aviation operations is currently limited.

Qualitative analysis of fatigue-related MORs nevertheless appeared to reveal considerable overlap in themes, particularly highlighting schedule-driven fatigue risks as major factors influencing the reporter's fatigue level, or considered by the reporter as the predominant safety concern in many MORs. At the aggregate level, a more systematic appraisal of reports mentioning schedule-driven fatigue risks of 'long duty days', 'late to early' shifts and lack of 'proper rest breaks' as drivers of on-duty fatigue requires the objective data that relate to those statements. It is notable that of the small proportions of MORs that had specified their duty or individual sleep timings, there were data examples indicative of extreme elevated fatigue risk similar to the Air Canada serious incident in 2017, such as the pilot having been awake and on duty over time periods between 17-19 hours (NTSB, 2017). Due to the limitations of the current MOR dataset, these specific findings should be interpreted with some caution. However, such data suggests the importance of more sophisticated bio-mathematical modelling analysis of fatigue-related risk concerns, since they may relate to probabilistic *predictable* fatigue risks threatening safety assumptions for pilot performance that are not unique to a particular set of events or conditions. Quantitative data capture of roster information furthermore would allow the regulator to establish whether such fatigue risks were predicted prior to flight based on the planned duty hours themselves or if these risks represented a deviation from the expected risks that occurred 'on the day'. Such insights relating to foreseeability of elevated fatigue risk matter in terms of ascertaining which organisational processes, rules or operational circumstances permitted elevated fatigue risks to occur, and whether these are unique (to a particular company or individual employee) or relate to broader industry working practices that reflect a common view on fatigue risk tolerance or highlight particular operational issues that have the potential to affect all airlines. Given the limitations of the MORs database, it is clear that future insights over distal fatigue risk factors would benefit from a much greater data capture and analysis of work schedule context and a common mode of reporting metrics to enable more precise cross comparisons between MORs. This would mean that the relationships between schedule-driven fatigue risks and safety occurrences could be more systematically investigated. This study also found that a large proportion of fatigue-related MORs carried no obvious operational consequences threatening the safety of the aircraft operation. However, this determination of severity of safety breach was limited by the fact that many reports lacked clarity on severity of sleepiness experienced, the timepoint and duration within the duty and whether the other flying pilot was also affected by fatigue.

In summary, fatigue is a prominent human factors hazard that has the ability to threaten pilot performance and flight safety. The ambition of safety occurrence reporting systems is to provide the basis for understanding the risk levels for hazards co-occurring or causal to the incident or safety event and provide evidence-based insights on suitable criteria for safety interventions or mitigations. Adopting a more systematic and thorough approach to fatigue data collection within incident reports is likely to dramatically improve the ability to assess and monitor fatigue risk trends over time, without the need to rely predominantly on word search terms to identify fatigue hazards. Moreover, in industries such as commercial aviation, which are supported by high reliability systems, it is important to discern both the contribution of fatigue to partial or complete performance failure of the pilot sub-system, and also to learn about the characteristics of incidents where a reduced safety buffer has in turn, decreased overall safety of flight. For this, reports of in-flight sleepiness with no obvious operational consequence should be considered important safety data for organisational learning. Such reports both represent ‘warning signals’ where distal fatigue-related defences have failed, and also have the potential to provide crucial insights on the factors and circumstances over why a more serious fatigue-related incident did not occur.

Chapter 4. Reported and predicted fatigue risk exposure of Pilot work hours

Survey and field studies conducted with commercial airline pilots suggest that in-flight sleepiness and related involuntary sleep phenomena are experienced by pilots during their duties. However, for methodological, practical and commercial reasons, there is a lack of publicly available research data of per-flight hour rates of sleepiness experienced by pilots or predicted fatigue risk rates associated with pilots' hours of work. Furthermore, the underreporting of severe on-duty fatigue via formal company or regulatory channels is a longstanding issue in aviation. Study 2 sought to address this gap by collecting self-reported sleepiness/alertness ratings from pilots from UK airlines via a mobile phone app over the period of August 2017. In tandem, predicted sleepiness levels and sleep lengths associated with participants' flown rosters were investigated using bio-mathematical fatigue modelling.

4.1 Ethics Statement

The study was approved by the University College London Research Ethics Committee, reference 8015/001. Participation was voluntary and informed consent was obtained by requiring prospective pilots to opt-in to the study via email, once they had read and were happy with the information sheet and study documentation. Participants did not receive any payment or reward for their time or effort.

Research Process

Figure 10 sets out the research stages taken in Study 2.

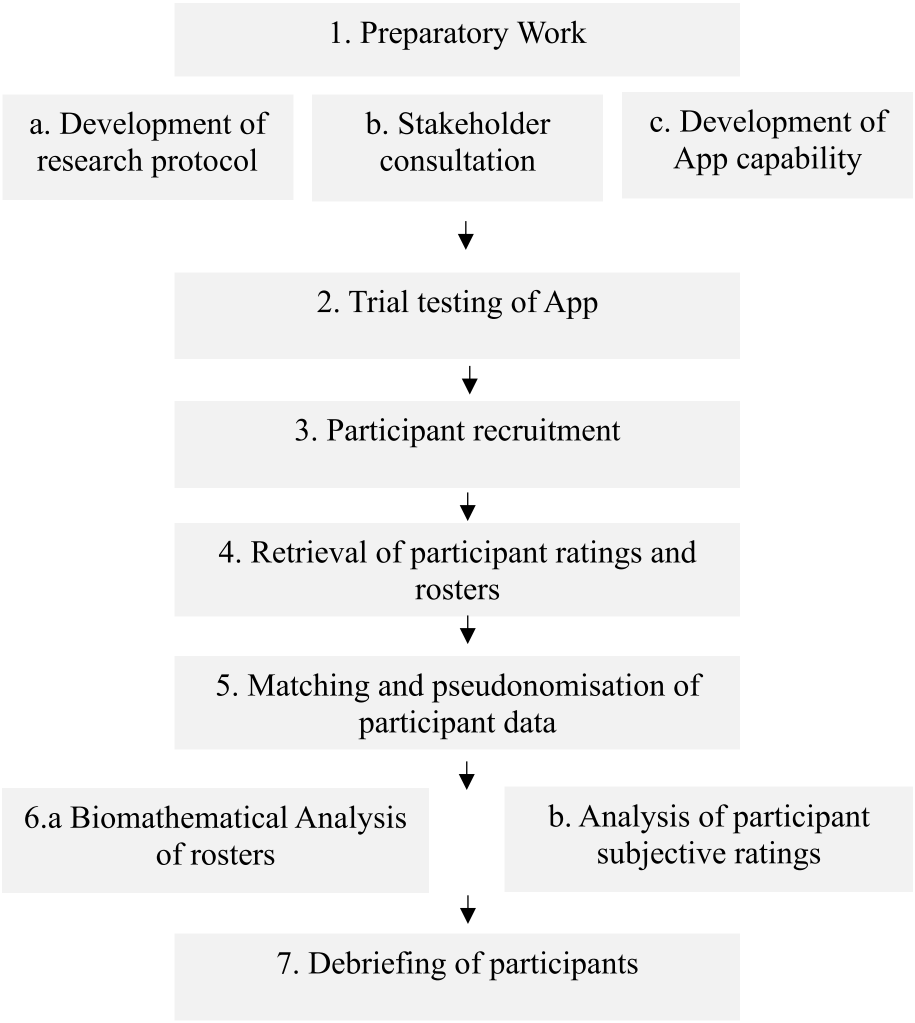


Figure 10. Research Process stages undertaken in Study 2.

Stage 1. Preparatory Work

Observational field research involving the collection of data from commercial pilots requires both thorough coordination between different stakeholder groups, and assessment of the range of considerations pertaining to the safety, ethical, legal, technical and practical issues associated with such research. Such considerations necessarily constrain the study design and research protocols. Hence, at the outset of this study process, several multi-disciplinary meetings with stakeholder groups were undertaken as preparation for the development and finalisation of a feasible study protocol that would enable the collection of sleepiness data and roster information from pilots whilst at work. This process involved consultations with operating and retired pilots, flight safety specialists, as well as BALPA flight safety sub-committee, legal and membership teams.

4.2 Development of the research protocol

The initial preparatory work for Study 2 involved a review of data collection opportunities for roster hours against different sleepiness or operator fatigue measures that could be used in field during a normal pilot work shift. The advantages and limitations of these measures were considered against the scientific literature (as set out earlier in Chapter 2, sections 2.3-4). An illustration of some of these measures set against in-field data collection opportunities is provided below in Figure 11.

The first objective of Study 2 was to retrieve pilot working hours to inform the shift driven fatigue risk exposure associated with UK airline pilot rosters (FRT 1). For this, pilot duty timing and location details were required as the substrate data needed for bio-mathematical modelling of fatigue risks for periods during flight, on-duty whilst not inflight, and estimated sleep opportunities prior to duty. Fortunately, the associated data retrieval mechanism for these data was relatively straightforward. UK commercial pilots are provided with a published roster both at the start of the month, and an updated version of this roster both prior to and post the completion of their flight duties. These roster documents contain the relevant duty time information (Duty Start and End times, Flight Times) alongside Airport codes (important for determining circadian shift, particularly in long-haul operations). This information is automatically recorded and updated in an electronic format (such as a PDF, .txt or .csv file) such that the pilot is able to send a copy of their ‘planned’ and ‘achieved roster’ (the actual timings of flights as flown) via email, with minimal time or effort incursion on their behalf. The one disadvantage associated with this retrieval method is that roster formats, including time zone information on publication (UTC, local base or other), whether report and debrief times were assumed (ie. times provided or principles such as ‘1 hour prior to first flight departure time’, ‘45 minutes post last landing’) are not consistent between different operators. The extent to which the formatting and content of rosters differed on such aspects was not known across UK operators prior to the study initiation but was logged as a project risk. Instructions to pilots on sending their rosters hence requested that where not visibly stated, the pilot should indicate how report and debrief times work in their company and to provide an explanation accompanying their roster. During the analysis period, where necessary these company principles were checked with pilot representatives from individual companies and two aviation scheduling specialists familiar with the company’s rostering processes.

The second objective of Study 2 was to ascertain pilot levels of sleepiness and sleep events that occur during their working day, as the primary data informing FRT 3 (behavioural symptoms and experience of fatigue during duty). Hence as a starting principle, methodologies that were inherently restricted in terms of the portion of day or operational location for which data collection could occur were decided against, since they would not provide in-flight information. These include cognitive tests which are sensitive to fatigue-related performance decrement at the time of testing, such as the psychomotor vigilance test (PVT). Such tests would not be operationally viable to undertake during normal flight and thus would provide data only to cover pre-flight and post-flight fatigue status.

Continuous measurement of pilot sleepiness levels during duty hours

As discussed earlier in section 2.3, continuous physiological measures of sleepiness tend to show greatest sensitivity to fluctuations in operator alertness and some (including EEG) have been used during normal flight operations. The data retrieval period for such measures could span from when pilots report in the morning to their debrief period at the close of their working day, providing a continuous profile of alertness measurements in pilots. However, such measures are typically expensive and labour-intensive, requiring substantial researcher involvement in setting up and individual tailoring of equipment, and may require the researcher to accompany flight crew during the observation period. As such, brain activity and ocular measures were considered to be impractical for use for a study involving pilots from multiple UK airlines. Ethnographic techniques involving direct observation and assessment of subjective or objective experience from pilots through systematic note taking of visible symptoms or in-situ researcher-led sampling of pilot fatigue experience through validated sleepiness scales were also considered. The major benefits of such an approach, commonly used in human factors research, is that the researcher has a greater ability to closely monitor subjects, retrieve a mixture of observational and rating data on different aspects of fatigue whilst the pilot is working, and form a richer appraisal of any subject-specific or environmental ‘on-the-day’ factors that may impact the participant’s level of alertness. In addition, by accompanying the participants throughout their workday, as illustrated in Figure 11, the available data capture period with which to appraise the subject’s alertness levels whilst at work is greater than other methods, and importantly can take place prior to, during and post flight. Despite these merits, for resourcing constraints and the ambition to derive a large sample size, such methods were not considered as appropriate for this study.

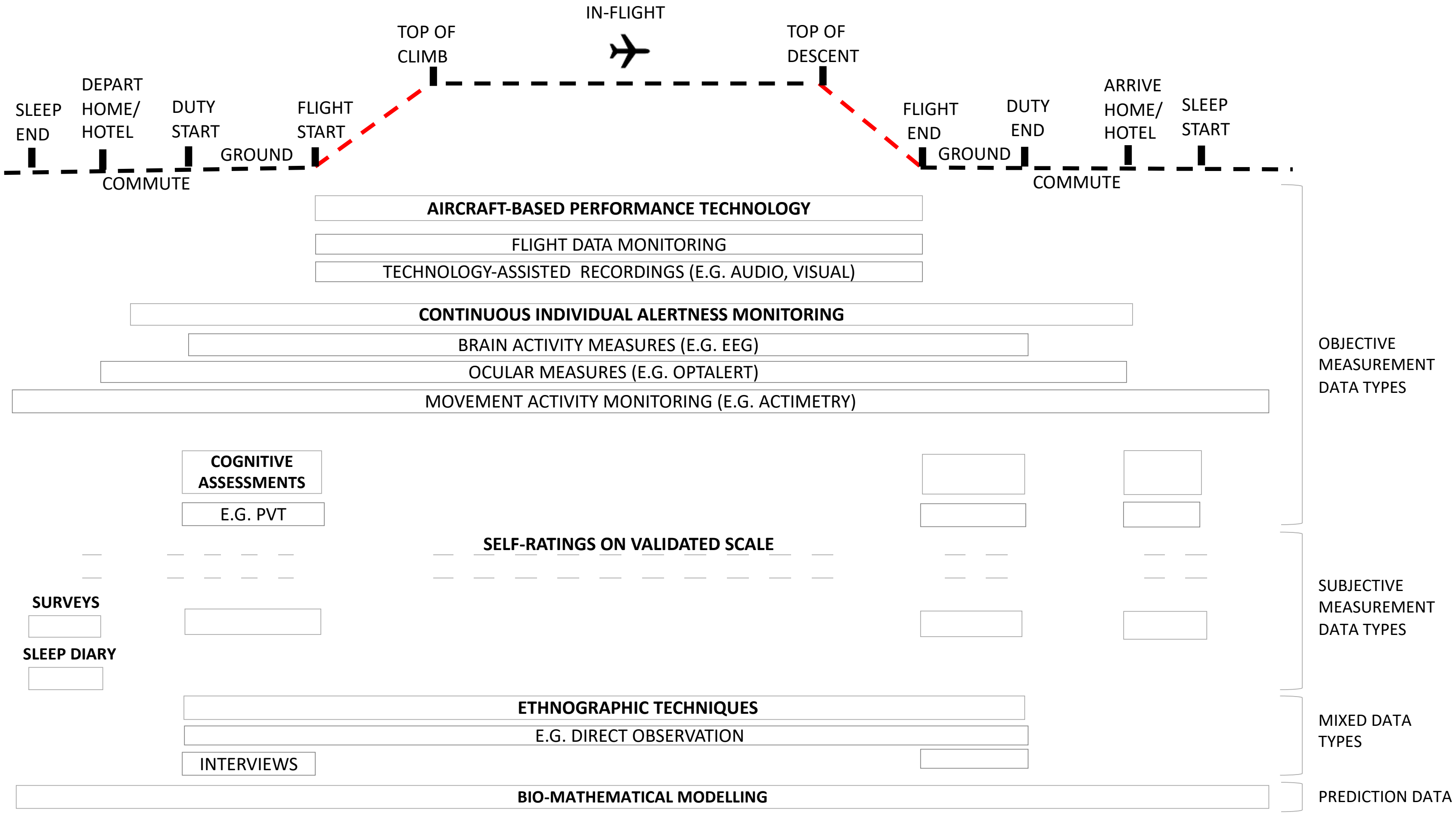


Figure 11. Data collection opportunities for different sleepiness or operator fatigue measures during a pilot work shift. Methodologies are grouped by measurement type and arranged in rows of varying horizontal length to illustrate their potential opportunity of use across the main operational and non-work timepoints during the pilot's day. For simplicity just one flight has been included within the duty day, however for short haul operations, (2-6) flights may be undertaken between the duty start and end times, which include additional ground times in between flight. The collection of Aircraft-based performance data was beyond the scope of this study, but is included for reference.

Subjective measurement of pilot sleepiness levels

Principle data collection methods of subjective experience of sleepiness include structured surveys or questionnaires which require participants to reflect on their sleepiness levels experienced prior to, during or post their duty day. The advantages of such methods are that these are low-cost and less resource intensive on behalf of the researcher, since they can typically be completed by participants on an electronic or paper form, without a researcher present. However, full survey or questionnaires which require a time incursion of for example, ten to fifteen minutes, necessarily must take place outside of the operational environment of flight or the on-duty ground period prior, where a pilot has reserved time to focus on their upcoming duties. From an operational standpoint, there is time during a pilot's working day in the debrief post-flight period to fill out a fatigue report, and as such this period represented a possible data collection point. However, it was decided that the research approach should not materially interfere with any operational period or process, particularly as pilots are obliged to fill out company fatigue reports if they experience severe fatigue that has impacted their performance during flight. Thus if fatigue symptoms had been present during flight, it was reasoned that the pilot participant should not have a dual commitment to fill out an operational safety fatigue report to their company alongside the completion of a research survey at the close of their day. It is conceivable that future research initiatives may be able to work closely with a company to align these legal and research processes into one. However, whilst important safety data, from the research perspective of sampling alertness across the duty day, this particular retrospective data collection method was not advanced further.

Subjective rating of sleepiness levels using the KSS validated scale involves participants selecting the number that best reflects their subjective experience just prior to rating from a simple Likert scale. As discussed earlier, the convergent validity and reliability of the Karolinska Sleepiness Scale (KSS) in measuring subjective sleepiness levels has led to its wider use in simulator and applied field settings. A major benefit is that ratings can be completed by participants quickly, easily and repeatedly across their working and non-working hours without a researcher present. Given this combination of benefits, self-rating on the KSS was advanced as the favoured measure for sampling pilot alertness levels across their duty day. However, the data collection method for self- ratings was also considered to be an important aspect to consider for the research protocol, since the ease with which pilots could provide KSS scores during their work day was seen as a major limiting factor in terms of the frequency of submissions provided by participants. Whilst noting ratings down on paper forms was considered, the priority was to try to find an approach where pilots from different companies operating at different times and locations could submit KSS data electronically on a regular basis during the study period. Ratings that are recorded and stored electronically are hence preferable since they may be more rapidly transferred into a research database, and do not rely on the participant to send back hard copy documents to the researcher. Some researchers, when working closely with a single company, have previously used the aircraft's systems to log participant ratings. For example, Powell and colleagues asked pilots to enter fatigue ratings directly into the flight management system of an aircraft just prior to top of descent, through designing a system that prompts pilots to enter the fatigue scale rating (in this case, Samn-Perelli) on a special screen (Powell et al., 2011). This allowed Powell's team to routinely collect fatigue data from crews with aircraft fitted with this modification. Whilst attractive as a method, the setting up of a system to interact with the Flight management systems of different types of aircraft, or indeed retrieve any additional aircraft or audio data from the flight data monitoring systems of participants was beyond the scope of this study. Instead, the idea of developing an app to allow pilots to submit research data including their KSS level ratings on their mobile devices during the study period was advanced. This approach had substantial merit since it was envisaged that a very high proportion of pilots would have a personal mobile phone or tablet device that they would keep with them throughout their time on duty, and thus the available period for data capture was not restricted. An additional benefit was the pre-existence of a BALPA membership '2-Way' app that many pilots already had on their phone, and which already stored basic personal information (such as pilot name, position and company) and permitted the electronic submission of reports. Hence, it was decided that Study 2 required the development of the 2-Way app to facilitate pilot fatigue ratings and submissions related to this research.

4.3 Stakeholder Consultation over App development

The following section describes the research considerations on content, usability and accessibility of using an app for the collection of participant data in the commercial aviation environment. The summarised main concerns and outputs from this consultation process are set out in Table 9 and described further below.

Table 9: Consultation over App development

App feature	Stakeholder Considerations	Main research protocol developments
1. Use	**1.1 Ease of use** Must be low effort, simple to use, easy to navigate, option for details not captured in data fields **1.2. Time incursion** Must be limited, ideally less than a minute	- Limit data fields to bare minima - Ensure click button format for most data fields - Option for free text narrative if desired - Maximise auto-fill or pre-select options - Design submission or data capture burden to be 'less than a minute'
	1.3. User comprehension Must be easy to understand and not require specialist knowledge	- Emphasize rating at all levels of alertness (different to fatigue safety reporting in aviation) - Develop specific user guide for rating via the app - Provide dedicated email address for app issues
	1.4. Timing of submissions Will differ for every participant and needs to be determined by the pilot	- The decision to rate alertness/ sleepiness level will self-paced and up to the pilot's discretion over when they make a submission. - Study guidance provided for optimal number of ratings (6-8) ideally spread across their hours awake.
	1.5. Engagement Must encourage engagement, but not distraction	- App must facilitate repeated submissions by 'pre-navigation' to the right section with each use - No timed prompts or push notifications as may distract participant whilst at work and reduce safety
	1.6. Data capture and transfer Must not rely on immediate internet connection, must not be burdensome	- Ability within the app to make a time-stamped rating that does not depend on internet connection - Ability to auto-upload ratings when participant reconnects to internet at later time point
	1.7. Timing information of sleepiness ratings Must collect accurate time information at the point of submission	- Ability within the app to accurately record time of rating, time of submission and time of data transfer - Timing information must be collected in UTC facilitated with auto-fill suggestion
2. Content	**2.1. Confidentiality of submitted data** Study participation will depend on reassurances that data will not be transferred to company or personal details revealed	- Emphasis within information forms and participant recruitment literature that data will be de-identified and part of a research study independent to any company. - Only aggregate information shared and published
	2.2. Ratings must not replace safety reports Study ratings must not conflict with other reporting responsibilities	- Details emphasized clearly on information forms - Short 'persistent' advisory to appear every time the app is opened - One time comprehensive disclaimer (is shown the first time, and no information can be submitted until a box is ticked to indicate the user has read and understood the disclaimer.)
3. Access	**Must work across different platforms** Most pilots carry phones or ipads on flight deck. A small proportion do not.	- App must be iphone and android compatible - Paper forms provided to pilots who wish to take part with hand-written forms

Consultation Process: Considerations for the collection of participant data via an app

1. Use

1.1. Ease of use and time incursion of the task of submission

As highlighted in Table 9 there were a number of considerations raised with respect to the pilots' ease of submitting ratings during their duties, and the practicalities of when the app would be used as part of this research study. A major set of considerations surrounded the time incursion involved with making a submission via the app. Pilot representatives and experts on the BALPA safety sub-committee highlighted the need to reduce the burden of this task to a bare minimum for several reasons. In particular, for gathering repeated measurements spread across the day, the task load needed to be managed effectively against the other pilot's tasks as part of their work duties. Hence any task of data submission relating to the pilot's alertness level needed to represent a minimal time incursion of a suggested period of 'less than a minute' to complete, and minimal cognitive effort in terms of user input so that the task was not off-putting to participants to have to complete multiple times across their period of being awake.

Protocol development 1.

In light of these points, the app data fields needed to allow for the collection of the minimal amount of useful data necessary to meet the study's aims. Such data fields hence require simple formatting, a reduced number of easy to use 'click' button options rather than open-ended answer fields, whilst also providing optional flexibility for additional narrative details to be added if desired and time permitting.

1.2. User comprehension

In order to distinguish the research requirement of alertness/sleepiness submissions from other existing practices of fatigue reporting within industry, it was emphasized that pilots should be reminded to submit ratings at all levels of alertness rather than just when they were subjectively feeling severe sleepiness levels. In addition various stakeholder discussions highlighted the need to provide additional clear guidance on the use of the app and submission expectations, beyond standard information sheets provided once prospective participants had opted into the study. This was felt to be particularly important since pilots would be opting into the study remotely and would therefore not be formally guided or monitored by a researcher during the study.

Protocol development 2.

In light of these points, a specific user guide was developed to aid pilots in their understanding of *what*, *how* and *when* to submit ratings via the app, in addition to the main study information materials. In particular participants would need to be reminded to submit ratings at all levels of alertness, rather than just when subjectively experiencing severe sleepiness levels. In addition, a unique email address was created for participants to highlight any technical or process issues relating to use of the app during the study.

1.3. Timing of participant submissions

Various research studies have conducted in-flight subjective ratings of fatigue or sleepiness, although fewer numbers have provided published peer reviewed data. For example, Gander and colleagues (Gander et al., 2013; Gander et al., 2015) have collected multiple ratings using KSS scores and other subjective measures in flight at the top of descent. One UK night time cargo operator also regularly collects pilots' sleepiness levels using Karolinska Scores at the top of descent via the in-flight computer (personal communication), and many more operators conduct similar exercises. The Federal Aviation Administration (FAA) and CAA (American and UK aviation regulators) have also been involved in similar research exercises during normal flights (CAA, 2003a). Hence there is industry-relevant precedent for the collection of subjective or objective sleepiness data from commercial airline pilots during normal operations. However, Study 2 uniquely sought to involve pilot participants from up to 21 airlines and was not coordinated with individual airlines or limited to specific routes or top of descent ratings. Hence at the outset, the study aims marked a departure from, and extension to previous research initiatives. As such, it was highlighted by various stakeholders that it would not be practical to mandate a fixed central coordination of timing of submissions from pilots participating in the study. In addition, from a safety perspective it was emphasized that pilots would be best placed to ascertain the times within their hours awake either on duty or not, within which they could make a submission. The rationale is as follows; in commercial aviation operations, during flight pilots have a mixture of prescribed 'SOPs' (standard operating procedures) that govern their actions during critical phases of flight such as take offs and landings; and general 'aviating', 'navigating' and 'communicating' tasks that couple their monitoring and interactions with the aircraft flight deck, their co-pilot and air traffic control. In a typical flight, there is a step change in the workload of flying the airplane that occurs at 20,000 ft (where usually the seat-belt sign switches on or off). Pilots have the expression that there is a ‘sterile’ or ‘silent’ cockpit below 20,000ft in take off or in landing, in which they avoid any unnecessary communication or subsidiary tasks. Outside of these critical phases, pilots will generally monitor the aircraft computer systems, and also fill out ‘tech’ logs, and review flight plans on their ipads, the inflight computer or on paper. Pilots are trained to manage and effectively prioritise tasks, and such, will only undertake these subsidiary technology log tasks when it is safe to do so and the workload is low. Within flight, it is after take off, during the cruise phase and prior to landing where it was determined pilots should fill out their Karolinska scale ratings.

Protocol development 3

Despite some clear methodological advantages of mandating consistency and adherence to specific data collection times intervals (either by time of day or operationally-defined points within a pilot’s duty) across the study population, these were not felt to be outweighed by the robust safety considerations and preferences of the end-user group pilot representatives. Hence the initial ideas over the study protocol needed to be updated to allow pilots to use their discretion over when to rate their alertness/ sleepiness level on work or non-work days. However, study guidance was also provided to give a suggested optimal number of ratings (6-8) to be submitted per day during the month long period, and the recommendation that such ratings would ideally be spread across the participants’ hours awake.

1.4. Continual participant engagement

A broad set of points were highlighted over the participant uptake and continued use of the App across the proposed month long study period. It was suggested that participant submissions may benefit from design features within the app to make repeated submissions easier once the app had been ‘opened’ once during a given day or night period. In addition it was highlighted that normal methods encouraging user engagement with an App such as repeated ‘push’ notifications (small messages that appear automatically as prompts on the user’s screen) would be inappropriate in a safety-critical environment, since they may cause distraction to the pilot whilst undertaking their duties. From a technical perspective such push notification methods may furthermore not always be possible during periods without internet, unless they adhere to pre-defined time periods.

Protocol development 4

In light of these points, the app was developed to facilitate repeat submissions by the addition of a quick navigate button to enable the opening of the fatigue rating and study submission section. The app was not set up to provide timed prompts or push notifications since these might distract the participant whilst at work and could reduce safety.

1.5. Accurate time recording of sleepiness ratings

There were various technical discussions with app and information technology experts pertaining to accurate recording of time information of the participants' submissions. These discussions highlighted the concern that aviation operations relate to 24/7 variable shift patterns, frequently involving time zone changes. As such, the recording of the time at the point of pilot submission must be consistently recorded, and the pilot must be guided in this process. In addition, there should be clear records of both the time of participant submission and the time of the data transfer. A clear example of this would be where a pilot would rate their subjective level of sleepiness in-flight via the app, but only achieve an internet connection for the transfer of that submission to the cloud data collection repository later that day or night. Use of any auto-timing features to capture time information within the mobile or tablet device would hence need to be assessed for inconsistencies or inaccuracies.

Protocol development 5

The App needed to be developed to accurately record three sets of timing information. First the time of rating, where the pilot makes a selection of their subjective alertness level within the app. Second, the time of submission (presumed to be approximately only less than a minute later) where the following clicks are made to submit the data information. Third, the time of the data transfer, where the full data submission is uploaded to the study cloud repository. The separate collection of these three sets of timings formed the basis of a series of systematic checks for trial testing of the app, to ensure accurate timing data would be collected and stored under varying permutations. Such checks involved submissions with and without internet access. The App hence additionally required a set of automated features; these included the need to prompt the user to confirm if the highlighted time on opening the app matched their understanding of the time in UTC. The app would also need to automatically upload any participant submissions when it was running and connected to the internet, without user input.

2. Content

2.1. Confidentiality of submitted data

Flight crew representatives suggested that for many in the pilot population, participation in a study collecting data on their on-duty alertness levels would critically depend on assurances over data confidentiality, and the independence of this research from their employer. In particular it was highlighted that many, if not most pilots would not want their personal data to be passed onto their employer or for the study to be run jointly with airlines. Lack of trust in the safety culture of their organisation and the perception of the potential for professional adverse repercussions to admitting fatigue whilst on duty were highlighted as two major issues that would dramatically disincentivise participation. These sentiments surrounding the underreporting of fatigue within commercial aviation are not new, and support the findings from a number of confidential surveys assessing pilot fatigue reporting in both the UK (ComRes, 2013) and across Europe (European Cockpit Association, 2012).

Protocol development 6

Participants would need to be assured that their data would be stored confidentially, and no personally identifiable data would be shared with their airlines. Furthermore, reassurances should be provided that only aggregate data insights would be shared and published. In addition, the collection of data would need to not involve company formal reporting systems. A password protected file would contain the participants name and their participant code to facilitate participant withdrawal if requested.

2.2. Participant’s reporting responsibilities

In order to distinguish the pilots’ pilot alertness/sleepiness submissions for the present study from existing safety practices of fatigue reporting within commercial aviation, it was emphasized on legal and safety grounds that pilots should be reminded of their professional reporting responsibilities to report any fatigue-related safety occurrences occur during the study period via formal channels. However, the European just culture regulation on reporting (Regulation (EU) No 376/2014) does not contain anything which prevents the monitoring of fatigue by researchers, so long as pilots continue to report mandatorily reportable occurrences to their airlines or the CAA.

Protocol development 7

The initial BALPA app instructions were designed to make it explicitly clear that pilots' KSS ratings would not replace the obligation to report any in-flight safety occurrences related to fatigue that may have compromised the safety of the flight to their company. Hence the app would need to include a short 'persistent' advisory to appear every time the app is newly opened (opened afresh) to alert pilots that their submissions via the app did not constitute formal safety data. The message should read;

'I understand that recording fatigue through this app complements my duty to report fatigue to my airline.'

The app should also include a one time comprehensive disclaimer outlining legal duties further (which can be shown the first time, and no information can be submitted until a box is ticked to indicate the user has read and understood the disclaimer). Subsequent uses will have the box ticked by default, and the disclaimer will not be seen.

3. Accessibility

Whilst most pilots carry phones or tablet devices on the flight deck, a small proportion do not or may not like to use their device for this purpose. Hence in order to make participation in the study accessible to individuals in the latter group, other methods of recording and submitting data should be considered.

Protocol development 8

The study should aim to cater for any pilots who wish to take part without use of electronic device. Hence a pen and paper format of capturing the participants' data should be developed and used instead.

The following section describes the 2-Way App and its developed data collection ability that was developed for Study 2.

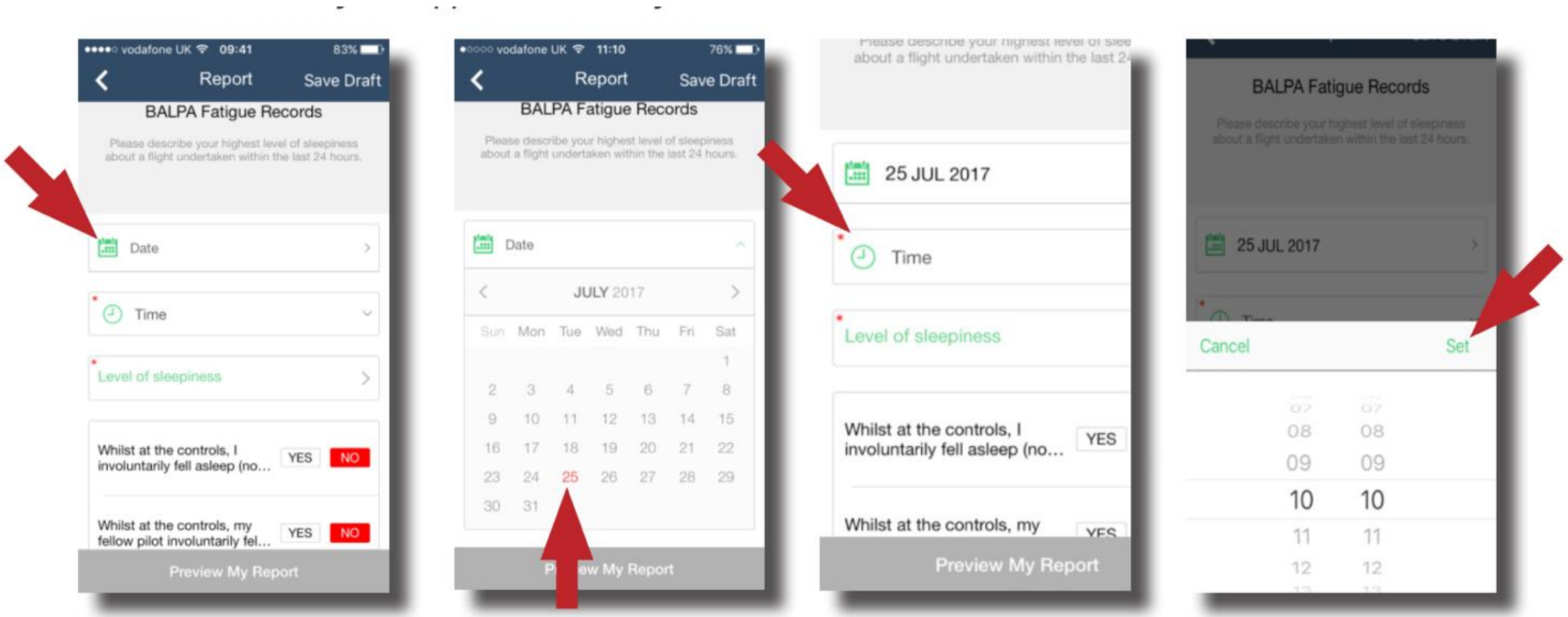
4.4 The BALPA 2-Way App

The 2-Way app is a free existing mobile / tablet device application, which enables BALPA pilot members to communicate with BALPA via specific report forms. As result of the stakeholder consultations during the preparation work for this research, a section of the '2-way' app was developed to enable pilots to submit data for research during the study period. The underpinning coding and software development for the Apple and Android interface on the BALPA-2-way App was carried out by BALPA's IT and software provider. The specification of all data fields of this app, guidance for use and collation and analysis of output data from the app were specifically developed and conducted as part of the research process for this thesis.

Figure 12a. The BALPA 2-way App icon

In addition to the KSS rating, additional timing information was required in order to time and date-stamp each individual rating submitted. An additional data point for collection in conjunction with this information was for pilots to select if they believed, at any point during their flight, they or their fellow flight crew member had involuntarily fallen asleep. In addition, there was an option for the participant to provide narrative information in a free text box, should they decide to.

The App was designed as follows: The user first needed to select the ‘Date’ button, where the current date is automatically highlighted in red in a drop down menu. If this is concurs with the correct date according to the user, the user clicks the red date, or if not would select the correct date from the menu. The user would then be directed to select the ‘Time’ tab, where current UTC time is set as the automatic default. This means that even if the local time is not UTC, the app should automatically set the counter to the UTC time. The user then clicks ‘Set’ if they are happy this automatic function has worked correctly and this populates the ‘Time field’.



BALPA Fatigue Records

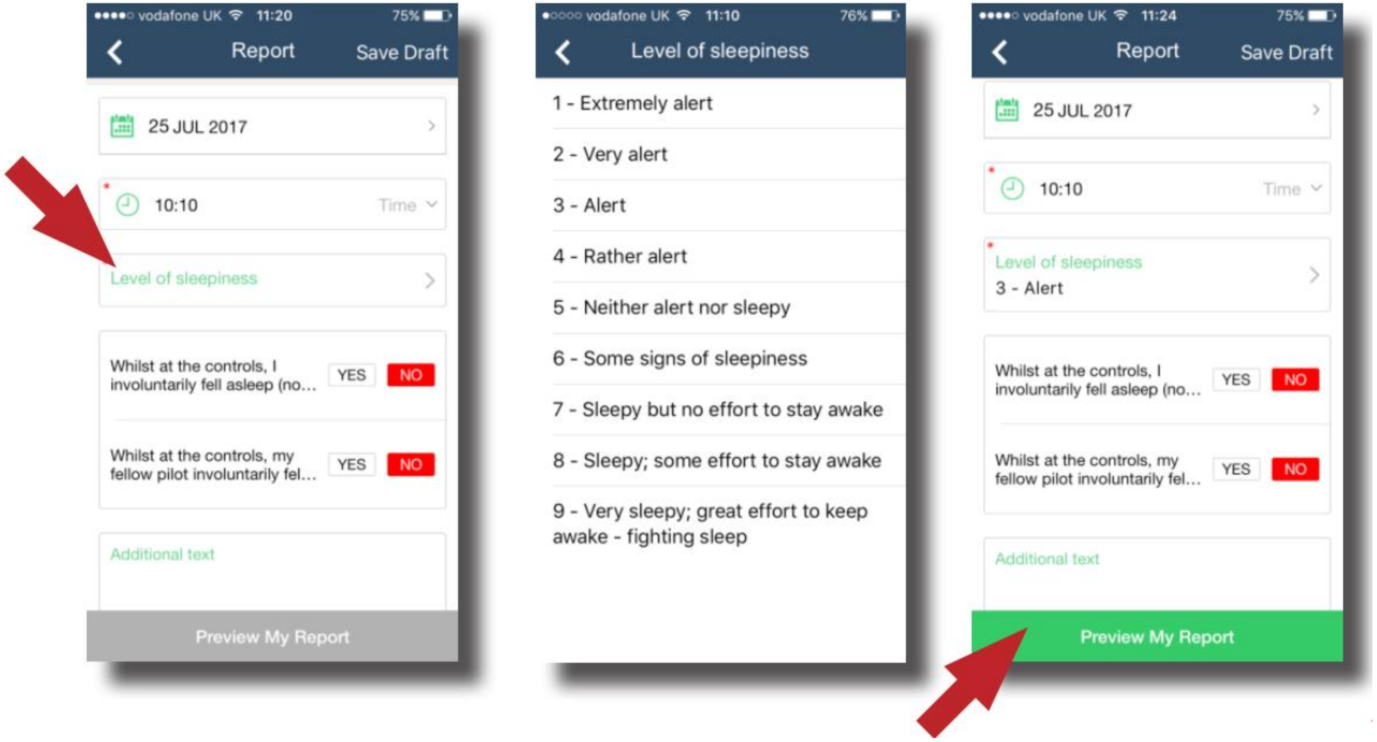
Please describe your highest level of sleepiness about a flight undertaken within the last 24 hours.

Whilst at the controls, I involuntarily fell asleep (no… YES NO

Whilst at the controls, my fellow pilot involuntarily fel… YES NO

Figure 12b. Screenshots from the BALPA 2-way App

The user is then asked to select the level of alertness or sleepiness that best describes how they feel, using the KSS scale. The Study information sheet advised that participants could submit ratings on both work days and non-work days. As such the following questions pertaining to if involuntary sleep had occurred at any point during flight optional to fill out, and were instructed to not filled be out on non-work days.



25 JUL 2017

10:10 Time

Level of sleepiness

Whilst at the controls, I involuntarily fell asleep (no…

Whilst at the controls, my fellow pilot involuntarily fel…

Additional text

1 - Extremely alert
2 - Very alert
3 - Alert
4 - Rather alert
5 - Neither alert nor sleepy
6 - Some signs of sleepiness
7 - Sleepy but no effort to stay awake
8 - Sleepy; some effort to stay awake
9 - Very sleepy; great effort to keep awake - fighting sleep

25 JUL 2017

10:10 Time

Level of sleepiness
3 - Alert

Whilst at the controls, I involuntarily fell asleep (no…

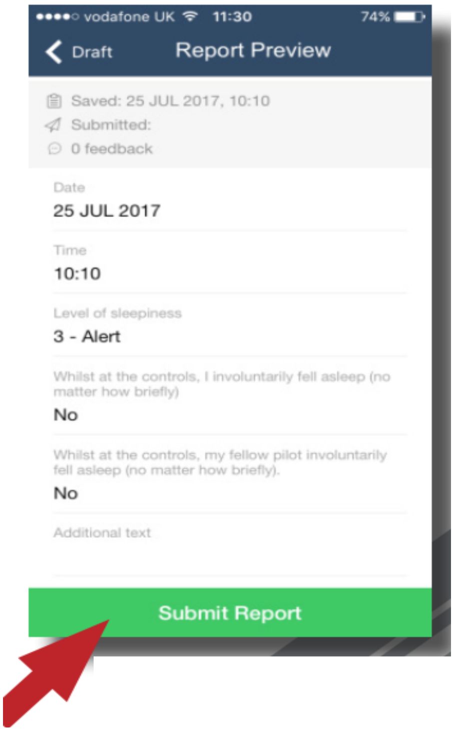
Whilst at the controls, my fellow pilot involuntarily fel…

Additional text

Figure 12c. Screenshots from the BALPA 2-way App

The statements below the sleep rating tab relate to the pilot’s previous flight. They are default set to ‘No’ in red. However, if during the study the participant or their fellow flight crew had involuntarily fell asleep whilst at the controls, they were instructed to select ‘Yes’ where the colour of this tab would change from white to red.

Pilots were instructed that their roster information coupled with the time information they provided would enable the researcher to determine which flight the involuntary sleep related to. There was an 'additional text' box available at the bottom of the screen for users to populate if there were further details they wanted to provide, but instructions for the Study emphasized that this was optional to fill out. Finally the user was instructed to click 'Preview my report', and then if content with the details to click the 'Submit report' button that appeared as an option on the preview page.



Saved: 25 JUL 2017, 10:10
Submitted:
0 feedback

Date
25 JUL 2017

Time
10:10

Level of sleepiness
3 - Alert

Whilst at the controls, I involuntarily fell asleep (no matter how briefly)
No

Whilst at the controls, my fellow pilot involuntarily fell asleep (no matter how briefly).
No

Additional text

Figure 12d. Screenshot from the BALPA 2-way App

The user could fill out and save reports whilst in flight mode on their device. Where this occurred, the app was designed such that the next time the user was connected to Wi-Fi and the app open on their device, their report would automatically upload.

Trial testing of the app

One month prior to launching the app for Study 2, the App was tested with a small sample (n = 11) of volunteers. The sample comprised of eight BALPA pilot reps and three Flight Safety staff members; sex (male = 10; female = 1), ages (25- 52), who owned a mix of iphone (n= 9) and android devices (n = 2). During the trial period these volunteers were asked to use the app as described in the information sheet. They were also asked to additionally make a separate note of the UTC times of rating. These timings were subsequently checked against the recorded information that had been transferred via the cloud to the study database. The volunteers were additionally asked for their feedback on the following questions

1) Is the time incursion for making a simple KSS time-stamped rating sixty seconds or less?
2) Does the app record ratings made when the device is both 'flight mode' and non-flight mode?
3) Are there any bugs or glitches with the app interfering with your ability to submit reports?

The feedback from volunteers was generally positive, indicating that the app was functioning as it should and was easy to use for this intended purpose. There were some display issues with the android version of the app in terms of the size of the dropdown display of KSS ratings and the partial occlusion of the main screen, which were quickly fixed by the app graphic design developer. The trial also confirmed that the rating timings that had been transferred to the study database concurred with the volunteer's notes of when they submitted reports, providing confidence that times were captured accurately.

4.5 Study Procedure

Participant Recruitment

Participants were recruited via the British Airline Pilots' Association (BALPA) membership database, which represents approximately 85% of all UK commercial airline pilots. An invitation to take part in the study was sent to eligible BALPA full members. Exclusion criteria included pilots who were retired members or not currently employed, or to those who have unsubscribed from membership communications, who were not contacted.

The BALPA 2-way app: Subjective sleepiness ratings and involuntary sleep reports

Pilots who agreed to participate in this study were asked to download the BALPA-2-way app onto their phone or tablet device, and instructed to submit six or more KSS ratings per day, both during flying and non-flying duty days, for a one month period. Participants were asked to rate at all levels of alertness (i.e. to submit KSS ratings not only when they were feeling sleepy, but also when they were feeling more alert). Participants were encouraged to spread their subjective ratings across their hours of wakefulness, although exact timing of the ratings were necessarily determined by the individual pilot, depending on their waking and working hours. In addition to the KSS rating, participants were also instructed to indicate any instances of involuntary sleep during flying duty days, where they and/or their accompanying flight crew had involuntarily fallen asleep at any point during flight. Participants were told that the free text box to add any operational or additional sleepiness details was optional to fill out. Full recordings containing the participants' rating, whether or not involuntary sleep occurred during the flight and free text submissions were automatically transferred to a database when the participant had an internet connection on their device.

Demographics

Participants were asked to provide standard demographic details (age, sex) as well as flight related details (role, flight experience in years) and additional sleep-related information (commute time and type, and subjective assessment of chronotype) via an online questionnaire at the start of the study period. For the assessment of chronotype, participants were asked to classify themselves as one of the following "very early", "early", "neutral", "late", "very late" ("One hears about morning and evening types of people, which one of these types do you consider yourself to be?") (Horne & Ostberg, 1976; Loureiro & Garcia-Marques, 2015). Prior research has suggested that individuals can reliably assess their chronotype through this single self-awareness item, and the score loads significantly together with all other morningness-eveningness dimension items (rMEQ; Horne & Ostberg, 1976), and shows convergent validity with the scale total score (Loureiro & Marques 2015).

Roster information and modelling assumptions

At the end of the August study period month, participants were asked to send in their flown rosters (achieved work shift patterns) for the months of July and August. The July month was retrieved to enable more accurate biomathematical fatigue predictions to be made for the first week of August. This roster information (including specific sector flights, duty periods, airport destinations) was then inputted into the Sleep-wake Predictor (SWP©: version 3.12; Åkerstedt et al., 2008) biomathematical model, for the computation of KSS predictions during work and waking hours, and prediction of assumed sleep lengths and timings. Rosters were entered into the model in the local time zone of home base, which for all participants was British Summer Time (BST). Key schedule information such as positioning duties (where pilots are travelling to their next flight within the duty period); standby duties (where pilots need to be prepared to be called for duty) and flying duties were further manually extracted from the rosters. Airport destination data (IATA codes) extracted from rosters were referenced against the corresponding time zone information obtained from the OpenFlights database for the purposes of working out time zone changes (OpenFlights, 2017). All other forms of activity (e.g. flight simulator training, ground training) were categorised as ground duties. Exact check-in and debrief duty times were inputted into SWP when this information was available in the rosters. For the minority of rosters which contained some, but not all of these full details, it was assumed that check-in prior to the flight time on rosters took place 1 hour prior (as the industry standard) and debrief period (the time from flight end till the end of duty period) was 30 minutes. Commute time was assumed to be 60 minutes as this is common practice when biomathematical models are used in commercial aviation, and the default assumption in SWP. In addition, a majority of volunteers (61%) indicated that their usual commute duration was between 30-90 minutes. For long haul routes, a number of variables including company, time of day, length of route and destination determine whether or not the number of flight crew is two or more. For example, in some companies three pilots are provided for single flights over 9-10 hours, whereas in others this is not the case. Given the large number of variables which differ between long haul participants' rosters, a decision was made to favour fewer assumptions and thus not apportion additional sleep opportunities for long haul rosters, in the case that there may be additional flight crew. In the present study, it is estimated that this assumption could affect 344.58 flying hours, which as a proportion of long-haul hours is 20% and 5% of the entire flying hour sample. This estimation is derived by calculating the number of epochs in the data where the pilots' roster indicates that they have been flying for 8 or more hours continuously, which may or may not be associated with in-flight rest opportunities, depending on the route, the time of day and the company the pilot works for. It was reasoned that any attempt to assume availability of in-flight rest would be a greater source of imprecision, affecting a greater proportion of the sample, than the decision not to make such an assumption. Four helicopter pilots also opted into the study, with two providing schedule and rating information, but due to their small number and inherent difference in operation from fixed wing operations, these participants' data were excluded from the Study 2 analyses. For the analysis phase, roster information was linked with rating and involuntary sleep data from the 2-Way App using the pilot's name and identification (ID) number as the reference point. Pilot names were subsequently de-identified and a new pseudo ID number created following this initial process. For each pilot, their roster information was collated in a single excel file that contained timestamped ratings and their predicted KSS scores and sleep opportunities prior to duty based on the SWP modelling from their roster information. These analysed rosters were drawn upon to provide the flying and duty hours per participant, alongside a number of other roster-related features not used by SWP. This included for example, the number of consecutive early or late starts during the pilot's flying duties and duty times that were shifted three or more hours prior to the previous duty start times. In addition, a master database was constructed that included aggregate information from all pilots and data sources as set out above.

4.6 Results

As seen in Figure 13, 294 pilots from 19 companies volunteered (3.12% of 9461 contacted members), and 95 pilots from 18 different airlines provided both roster and rating information. Pearson's chi-squared test of independence revealed that the participant sample was not significantly different to the membership population in terms of company break down χ^2 (180) = 190, p=0.29), gender χ^2 (1) = 0.86, p = 0.35 or age, t (312.11) = 1.93, p > 0.05. Further descriptive statistics of the pilot sample are provided in Table 10.

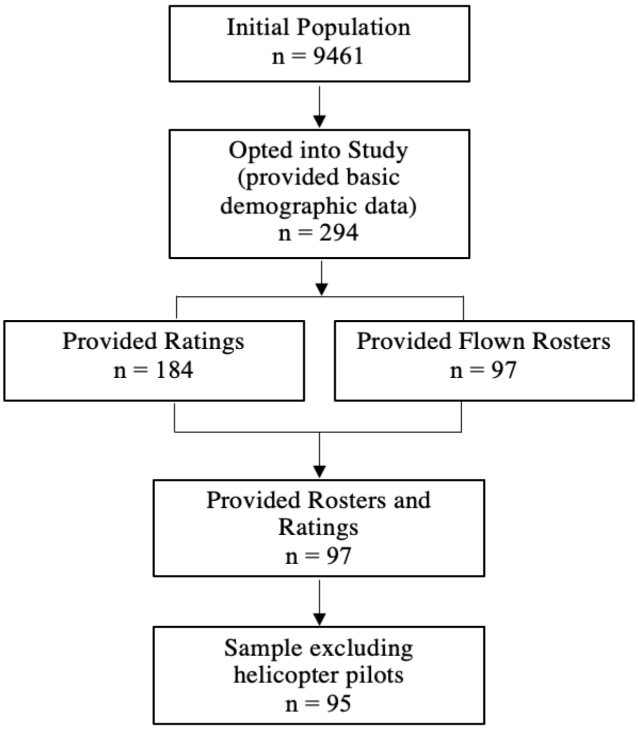


Figure 13. Flow chart indicating participant numbers for different categories of data

Demographics of study sample

Table 10 shows the demographic details of the pilot participants who opted into the study, and those participants who submitted both KSS ratings and complete rosters during the one month period.

Table 10. Descriptive statistics of participants

		Opt in n= 294		**Provided Rosters & Ratings n= 95**	
Sex					
	Male	272	92.5%	85	89.5%
	Female	22	7.5%	10	10.5%
Age					
	Years (*M* ± *S.D.)*	42.1 ± 9.7		42.0 ± 10.3	
Flight experience		**n= 175**			
	Years (*M* ± *S.D.)*	15.2 ± 9.0		15.6 ± 8.8	
Chronotype					
	Morning type	**n/a**		27	28.4%
	Intermediate type			14	14.7%
	Evening type			50	52.6%
	Missing data			4	4.2%
Role		**n= 292**			
	Captain	154	52.7%	47	50%
	First Officer	135	46.2%	47	49%
	Other	3	1%	1	1%
Operation type					
	Long haul	**n/a**		23	24.2%
	Short haul	**n/a**		72	75.8%

Table values are provided to one decimal place. n/a refers to data that were not available; *M* refers to mean and *S.D.* refers to standard deviation. Of participants that opted in, some provided incomplete data, so subset participant numbers are listed for flight experience and role.

Description of pilot working hours

Within commercial pilot work schedules, the term ‘Flying hours’ relates to the period of time between the aircraft being off blocks (when the aircraft becomes free to move) and on blocks (where the aircraft is restrained from moving). ‘Flight duty period’ (FDP) refers to a duty which includes flying time, the turnaround time, and one hour pre-flight preparation. A ‘Duty period’ may relate to a period of work which includes flying and/or non flying duties; where the duty period includes an FDP, the period will additionally include the debrief time at the end of a flight, which is approximately 30 minutes for most airlines[3]. Typically, long haul routes only include one sector, whereas short haul routes range from two to six sectors. Long haul operations are not normally followed by an immediate flight the next day, and so consecutive duty starts in this study almost exclusively refer to short haul operations. Table 11 shows flight and duty information from pilot’s work schedules. For the purposes of this study, ‘Super early start’ duties were defined as check in times between 00:00- 06:00 BST;

---

[3] An illustration of these terms against the pilot working day may be found in Appendix i.

‘Early start’ duties referred to check in times between 06:01- 09:00 BST; ‘Late finish’ duties referred to duty end times after 00:00 BST. Window of circadian low or ‘WOCL’ duties referred to duty starts between 00:00-06:00 BST or duty finishes between 0000-0800.

Table 11. Flight and duty data from pilot’s work schedules

Flight and duty information	Long-Haul	Short-Haul	Overall
Flight Period			
Average Flight length (hours) (*M*, *S.D.*)	9.4 ± 2.6	2.2 ± 1.1	
Number of Flights	180	2414	2594
Number of Flight Duty periods	172	932	1104
Flying Hours (total)	1692.6	5294.1	6986.7
Duty Period			
Average Duty Length (hours) (*M*, *S.D.*)	11.2 ± 3.1	8.5 ± 3.2	
Number of Duty Hours	2123.6	9374.3	11497.9
Number of Duty periods	190	1105	1295
Super Early Start Duties	43 (22.6%)	267 (24.2%)	310 (23.9%)
Early Start Duties	56 (29.5%)	533 (48.2%)	610 (47.1%)
Late Finish Duties	64 (33.7%)	115 (10.4%)	197 (15.2%)
WOCL Duties	106 (55.8%)	379 (34.3%)	485 (37.5%)
Consecutive duties starts before 06.00			
2 x Consecutive Super Early	1 (0.5%)	118 (10.7%)	119 (9.2%)
3 x Consecutive Super Early	0 (0%)	51 (4.6%)	51 (3.9%)
Consecutive duties starts before 0900			
2 x Consecutive Early	1 (0.5%)	326 (29.5%)	327 (25.3%)
3 x Consecutive Early	0 (0%)	188 (17%)	188 (14.5%)
Consecutive duties with late finishes			
2 x Consecutive Late Finish	1 (0.5%)	27 (2.4%)	28 (0.1%)
3 x Consecutive Late Finish	0 (0%)	3 (0.3%)	3 (0.2%)
Duty start times shifted (3 or more hours) relative to previous duty start times	32 (16.8%)	178 (16.1%)	210 (16.2%)

Within our sample of 1295 duty periods, duties that started or ended during the known WOCL periods made up over a third of all duties, with 310 (24%) having check in times before 0600 and 197 (15%) with duty end times after 00:00. For both long and short haul pilots, 16.8% of duty start times in this study were shifted 3 or more hours relative to previous duty start times.

Table 12 shows the biomathematical predictions of main sleep opportunities prior to flying duties. ‘FDP’ relates specifically to a duty period which includes flying duties. Sleep periods of less than three hours duration (typically pre-flight ‘nap’ opportunities ahead of late duties) were excluded for this table.

Table 12. Biomathematical Predictions of main sleep periods prior to flying duties

Flying Duty Periods (FDPs)	Long Haul	Short Haul	Overall
Predicted sleep length (M ± S.D.)	7.4 ± 1.0	6.8 ± 1.2	6.9 ± 1.2
Median predicted sleep length (IQR)	7.8 (7.5 - 7.8)	7 (5.8 - 7.8)	7.33 (6.0 - 7.83)
Predicted sleep period <7 hours	30 17.4%	441 47.3%	471 42.7%
Predicted sleep period <6 hours	16 9.3%	259 27.8%	275 24.9%
Preceded by three consecutive sleep periods <7 hours	0 0%	105 11.3%	105 9.5%
Preceded by three consecutive sleep periods <6 hours	0 0%	20 2.1%	20 1.8%

Welch’s t-test revealed that there was a significant difference between the mean predicted sleep lengths prior to flying duties between short and long haul work patterns t (281.4) = 7.73, $p < 0.001$, and examination of the interquartile ranges reveals the greater spread of predicted sleep opportunities in short haul (5.8-7.8 h) compared with long haul operations (7.5-7.8 h). When considering the main sleep opportunities prior to duty, a quarter (24.9%) of flying duty periods were predicted to be preceded by a sleep period of less than six hours, extending up to 42.7% predicted to be preceded by a sleep period of less than seven hours. Since there were greater numbers of short haul pilots who both opted into the study and provided full roster information (SH n= 72; LH n= 23), the overall proportion reflects the greater abundance of short haul duties within the analysis. With respect to repeated shortened sleep periods, over 10% of short haul flying duty periods were predicted to be preceded by three consecutive sleep periods of less than seven hours, and 2.1% less than six hours.

Within our dataset a small proportion of flying duty periods (32 out of the study sample of 1104 FDPs) were associated with a pre-flight sleep period of less than three hours. Most of this subset of predicted sleep opportunities less than three hours (28 FDPs) related to long haul duties where the model predicts an additional day sleep opportunity period prior to a late departure flight. For the purposes of establishing the mean predicted sleep length prior to flight across all duties, such sleep opportunities were assumed to represent an opportunity for a pre-duty nap and hence eliminated from the analysis in Table 12, since they did not represent the main sleep opportunity prior to flight and would disproportionately reduce the mean long haul pre-flight sleep lengths. However, for all following analyses pertaining to biomathematical model predictions concerning on-duty alertness, these predicted nap opportunities were retained.

Biomathematical Predictions of Karolinska Sleepiness Score (KSS) levels at Duty points

Table 13 shows that the mean predicted sleepiness scores at key operational points (duty start, last landing and duty end) were overall not indicative of severe KSS sleepiness scores (predicted KSS 7 or above), in long haul or short haul flights in this study, although the standard deviation values indicate a reasonably large degree of variation around these means, particularly in long haul rosters.

Table 13. Predicted KSS Scores at operational duty points

	Predicted KSS level (M, S.D.)	
Operation Point	**Long-Haul**	**Short-Haul**
Duty Start	5.06 ± 1.51	4.53 ± 0.85
Last Landing	6.57 ± 1.63	4.94 ± 1.05
Duty End	6.68 ± 1.68	4.96 ± 1.11

As may be viewed in Table 14, elevated KSS levels were predicted to occur during approximately 10% of flying hours, within which 225.7 hours (3%) were associated with predicted fatigue levels of KSS 8 or above. The majority of flying hours associated with elevated KSS predictions were from long haul schedules. 12% of flying hours in the sample were associated with continual hours of wakefulness in excess of 16 hours. On average short haul pilots were likely to have been awake for 11.56 hours at last landing, (*SD* = 3.76), compared with long haul pilots with an average 18.31 hours, with a large deviation around the mean (*SD* = 5.6).

Table 14. Flying hours associated with high KSS predictions and elevated continual hours of wakefulness

	KSS Prediction		**Continual Hours Wakefulness**		
	KSS >=7	KSS >= 8	>=16 hours	>=17 hours	>=18 hours
Long Haul flying hours	591.7	208.8	647.0	558.2	465.0
% long haul flying hours	35.0	12.3	38.2	33.0	27.5
Short Haul flying hours	95.0	16.8	201.7	101.7	45.5
% short haul flying hours	1.8	0.3	3.8	1.9	0.9
Total	686.7	225.7	848.7	659.8	510.5
% of total flying hours in sample	**9.8%**	**3.2%**	**12.2%**	**9.4%**	**7.3%**

Karolinska Score Ratings

8291 ratings were provided by participants during the study period via the 2-way-App, and 5382 ratings from the 95 pilots who submitted their full rosters, with this subset of participants submitting an average of 55.5 ratings across the one month period but showing considerable variation in submission rates (*M* = 55.5, *SD* = 44.6). There were 140 ratings submitted by 26 pilots who had not formally opted into the study or provided roster or demographic information. These data were not used in the study analyses. As shown in Table 15, the submission rate across the study period remained consistent, with only a slight drop in submissions towards the end, where some roster blocks extended into the following month and participants continued to rate until the end of such blocks.

Table 15. KSS submissions across the study period

Week	Number of Ratings
1	1376
2	1925
3	2011
4	1873
5	1106

Figures 14a and 14b show the relative distribution of ratings between operation type and time of day. The majority of ratings clustered around the middle values of the scale, suggesting that as instructed, in their waking hours pilots were providing ratings at all states of alertness/sleepiness, and not just when they felt sleepy. Within the sample there are a greater number of short haul pilots and as such the frequency of ratings submitted across the entire KSS scale appears to be concomitantly higher. Both short haul and long-haul pilots furthermore were submitting ratings across the 24-hour period, although with a reduced submission rate during the early morning hours (00.00-06.00).

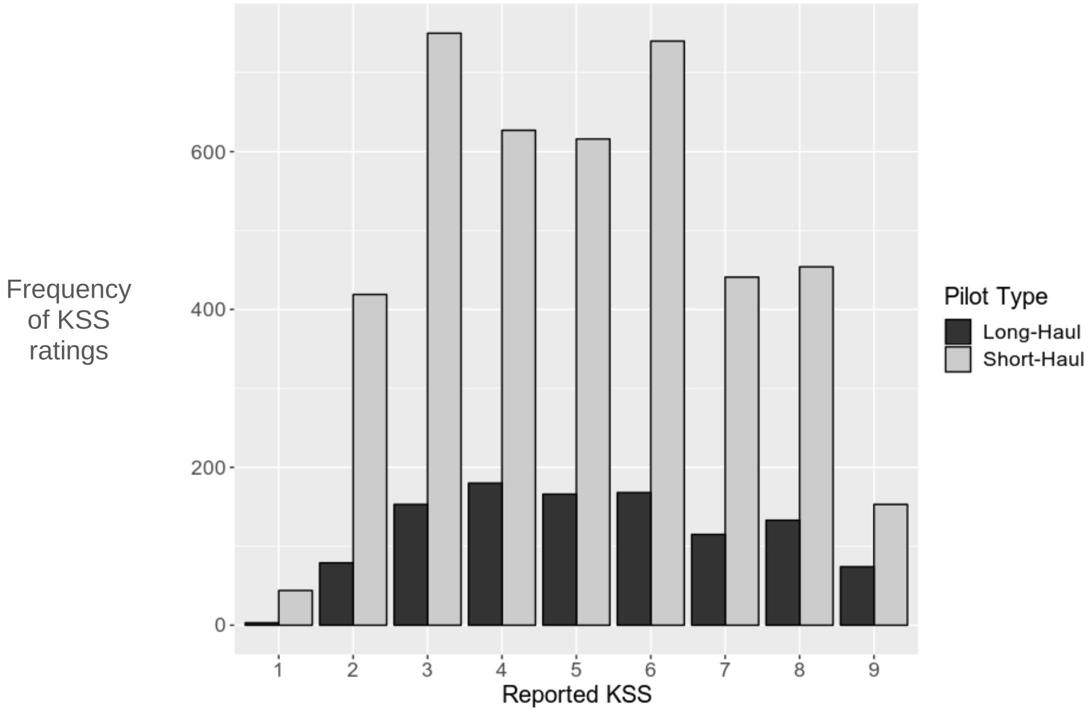


Figure 14a) Overall frequency of KSS ratings during flying and non-flying hours for long and short haul pilots.

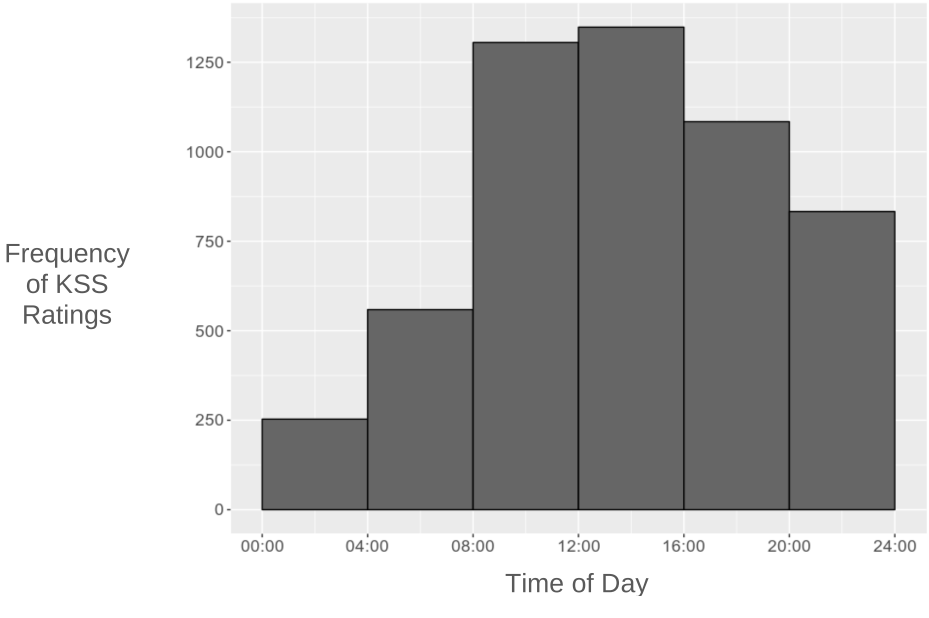


Figure 14b) Frequency of KSS ratings plotted in six 4-hour bins across the 24 hour period. Time of day is expressed in British Summer Time (BST).

Thirteen participants submitted ten or less ratings during the study period. This finding was investigated to see whether pilots with low submission rates had only provided ratings when feeling sleepy (like they would normally for fatigue reporting to their safety departments). Visual inspection of these participants' data suggested that in fact these participants showed a similar, albeit higher KSS rating pattern to the rest of the study sample (Figure 15). The average KSS rating for participants submitting ten or less ratings over the one month study period ($M = 6.6$, $SD = 1.3$) was found to be significantly greater to those who submitted greater than ten ratings ($M = 5.16$, $SD = 1.1$), $t(76) = 4.14$, $p < 0.001$.

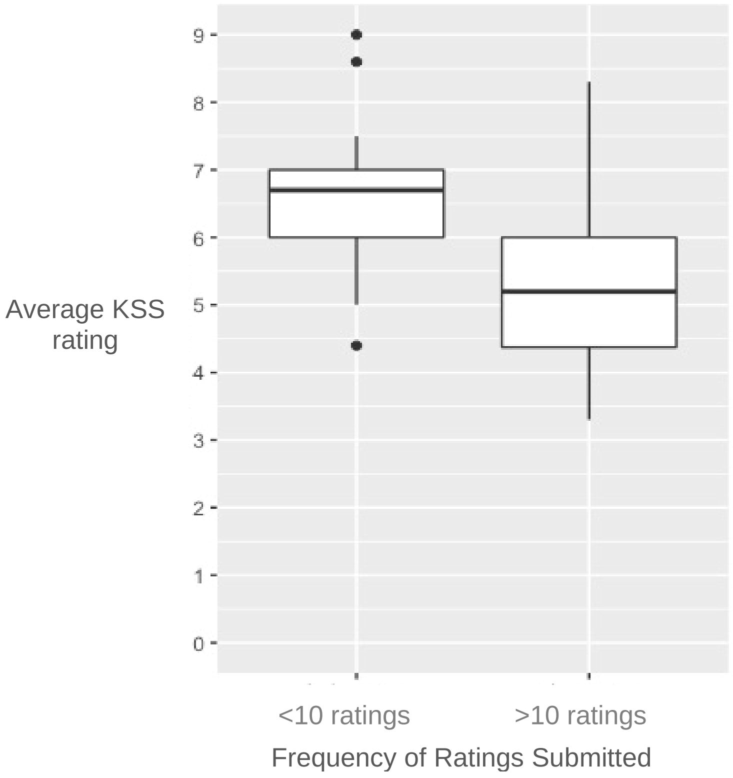


Figure 15. Box plot comparing interquartile range of average KSS ratings by total ratings submitted

In-Flight KSS Ratings

During flying hours, participants submitted 839 KSS ratings, which were distributed across the entire range of the KSS scale for pilots flying both long and short haul operations (see Figure 16a). At the higher end of the scale, there were 211 in-flight KSS ratings (148 from short haul pilots, 63 from long haul pilots) at or above KSS 7 during the one month August period, equating to an overall rate of 1 KSS rating of 7 and above per 33.1 flying hours. There were 87 KSS 8 ratings (30 from LH pilots, 57 from SH pilots) and 34 KSS 9 ratings (15 from LH, 19 from SH), which equated to an overall rate of 1 report at or above KSS 8 per 59 flying hours. As may be seen in Figure 16b), whilst both long and short haul pilots submitted high KSS scores during flight, long haul pilots submitted a disproportionately high number of KSS 8 and 9 ratings relative to short haul pilots, in line with biomathematical model predictions of the greater proportion of long haul flying hours associated with KSS 7 and above.

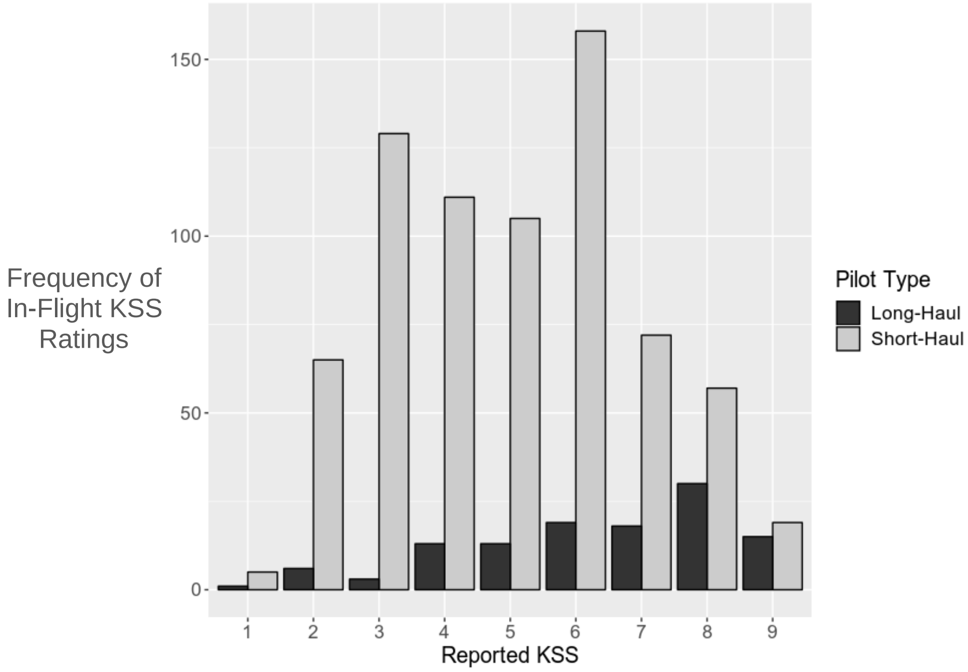


Figure 16a) Frequency of in-flight ratings by reported KSS

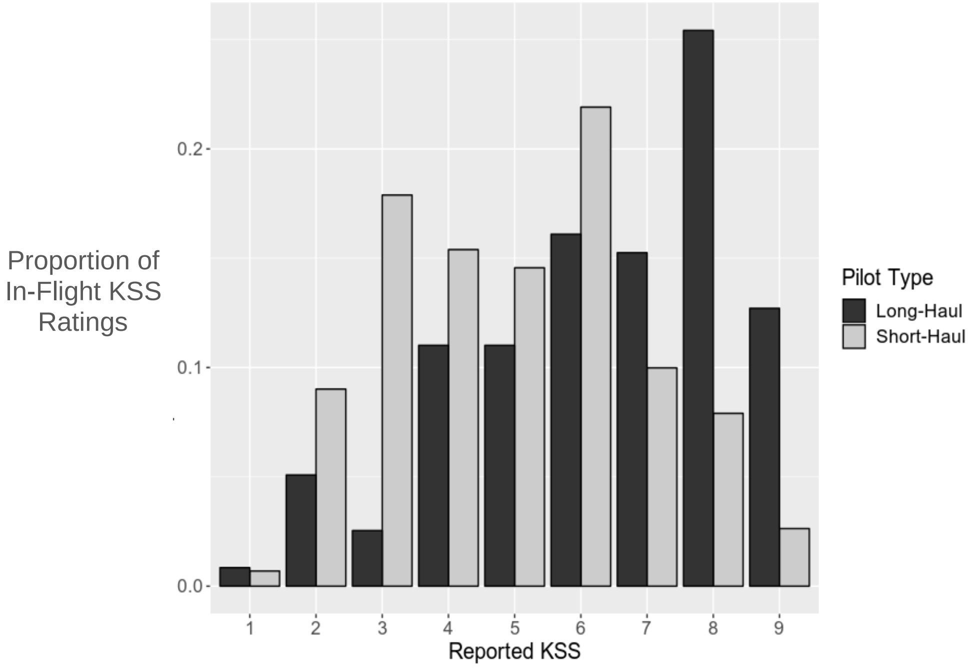


Figure 16b) In-flight KSS ratings expressed as a proportion of all submitted ratings by pilot type

Predicted and reported KSS

KSS rating data were further investigated to assess the relationship between predicted and reported KSS in this study. To begin with, as suggested by Field (Field et al., 2012) the data were analysed using simple linear modelling to provide a baseline for comparison. This analysis for 'Model 1' indicated a significant effect of predicted KSS on reported KSS levels, *F* (1, 5380) = 412.2, *p* <0.0001. The slope coefficient suggested that as predicted KSS scores increase by 1, reported KSS increases by 0.6. The Adjusted R^2 value of 0.071 however was low, suggesting that whilst there was a significant overall relationship, there appeared to be substantial variance in reported KSS levels that is not explained by predicted KSS levels. Although a useful starting point for analysis, simple linear regression analysis is not suitable for these data since the outcome variable (reported KSS) is a repeated measure and as such, use of simple linear modelling violates the assumption that observations are independent from each other. Therefore linear mixed model analyses were subsequently used to analyse the data.

The first linear mixed model considered was to use predicted KSS as a fixed variable and the intercepts modelled as random effects by the level of the individual participant ('Model 2'). This allowed the model to consider relationships between predicted and reported KSS at the individual level, and contain differing amounts of within-participant submitted data. This analysis suggested again an overall significant effect of predicted KSS on reported KSS *t* (22) = 0.6, *p* <0.0001, such that as predicted KSS increases by 1, reported KSS increases by 0.6. Interestingly, assessment of random effects for individual pilot participants suggested that all pilots had similar intercepts, since the variance was low and less than the standard deviation (*Var* = 0.97, *SD* = 0.98), indicating that different pilots did not substantially differ in terms of their reported KSS at the intercept. As such, this analysis suggested that on the whole, as predicted KSS scores increase, reported KSS increases for all pilots. A subsequent step was taken to run a model with predicted KSS as both a fixed and random effect ('Model 3'). Examination of the fixed effects revealed that the regression parameter for the effect of predicted KSS reduced slightly to *b*= 0.55 but remained significant *t* (64.52) = 10.85, *p* <0.0001. Examination of random effects revealed that the variance of the *intercept* had increased from Model 2 to Model 3 and exceeded the standard deviation. (*Var* = 4.34, *SD* = 2.08). Comparatively the variance in KSS as a *random effect* was limited and lower than standard deviation (*Var* = 0.14, *SD* = 0.37). The slopes and intercepts were also highly correlated *r*= -0.88. The outputs of Model 3 hence suggested that when the slopes are allowed to vary (with the addition of predicted KSS as a random effect), the variation in intercepts also increases. Model 3 was furthermore found to be a significantly better fit for the data compared with Model 2, χ^2 (2) = 76.6, $p<0.0001$. However, variation in the slope gradient and direction did not appear to differ substantially between pilots, and as such a similar relationship was observed between predicted and reported KSS across pilots.

Further exploratory investigation was undertaken to assess whether the addition of certain roster-related variables to Model 3 increased prediction accuracy of reported KSS. Whilst there are no established ways of calculating the correct sample size for mixed effects modelling, a suggested rule of thumb from statistical authors is 20 participants per predictor variable (Field, 2009; Miles & Field, 2012). Analysis was therefore limited to assessing predictors of theoretical interest in line with the sample size of 95. Particular roster-related variables of greatest theoretical interest for this study were included as fixed effects. These were; sleep hours in the last 24 hours, continual hours of wakefulness, number of consecutive early duties in the last 72 hours and the number of consecutive late duties in the last 72 hours. Age and chronotype were also initially investigated as fixed predictors in Model 4, but eliminated from further analysis due to non-significance. Table 16 provides a summary of the parameter estimates and standard errors of the predictor variables included in the models.

Table 16. Comparison of Models 3- 6 outputs

	Model 3		*Model 4*		*Model 5*		*Model 6*	
Fixed effects	*b*	SE	*b*	SE	*b*	SE	*b*	SE
Intercept	2.85***	0.27	1.88***	0.52	3.18***	0.38	2.92***	0.26
Predicted KSS	0.55***	0.05	0.55***	0.05	0.36***	0.06	0.37***	0.05
Age			0.02	0.01				
Chronotype			0.06	0.08				
Sleep hours					- 0.02	0.03		
Continual Hours of Wakefulness					0.08***	0.01	0.08***	0.01
Consecutive early duties					0.1*	0.04	0.11**	0.04
Consecutive late duties					0.29**	0.1	0.31**	0.1

Coefficient estimates and standard errors are provided to 2 decimal places. *p<0.05, **<0.01, ***p<0.0001

As shown in Model 4, the *b* coefficient for Age was low, and although close to significance, was not found to significantly predict reported KSS t (77.9) =1.97, $p = 0.051$, and chronotype was also non-significant t (79.1) = 0.82, $p = 0.82$. In Model 5, predictors that took into account the continual time of wakefulness at the time of the participant rating, and the number of consecutive early or late duties in the last 72 hours were found to be significant. The addition of these three variables resulted in a significantly better fit compared with Model 3, $\chi^2(4) = 209.5$, $p<0.0001$. Model 6 omitted sleep hours (as this was non-significant in Model 5) and was subsequently run with the significant predictor variables from Model 5. ANOVA analysis of models 3, 5 and 6 revealed that Model 6 provided the best fit for the data, and represented a significant improvement in model fit χ^2 (3) = 208.59, $p <0.0001$.

The data were also assessed to see whether the KSS predictions were more likely to underestimate or overestimate pilot reported KSS. Figure 17 shows participants' individual KSS ratings (black dots) submitted during the August month period, plotted against the level of KSS predicted from the participants' work schedule, for the point in time the rating was submitted.

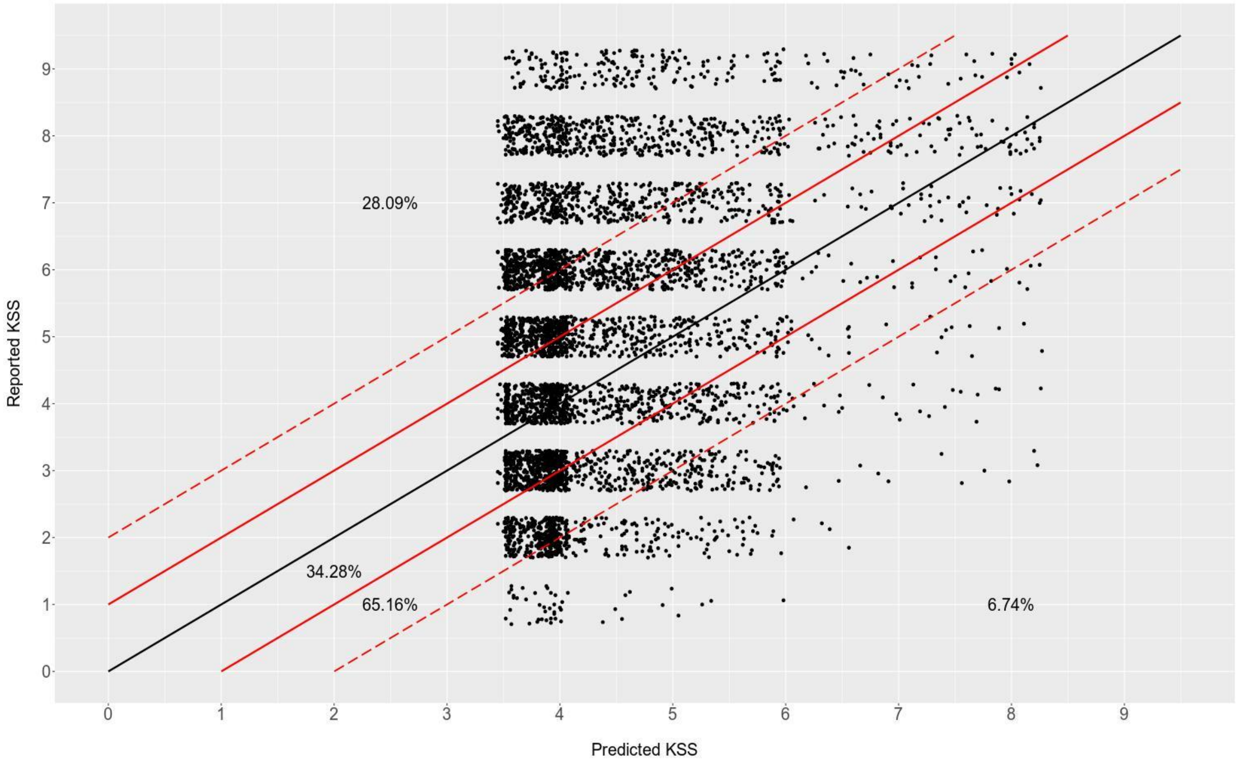


Figure 17. Predicted KSS and Reported KSS scatterplot
The black line is the line of equivalence between the predicted KSS scores and the reported scores, which indicates perfect agreement between model predictions and ratings. The solid red line denotes ratings that are within 1 KSS point above or below the KSS prediction at the time the rating was made. The dashed red line indicates ratings that are within 2 KSS points above or below the KSS prediction. To aid visualisation, the points have been randomly jittered by +-0.3 on the y-axis so that they are not overlapping, and density of points may be visualised.

Schedule-driven model predictions of KSS exist on a continuous scale, whereas self report ratings of KSS are provided on a 9 point discrete Likert scale, which means that there is a general bunching of points with respect to the y- axis of reported KSS level, but not the x-axis of KSS prediction level. So that the relative density of ratings may be seen, the points have been jittered along the reported KSS scale (y- axis) by a small degree. Schedule-driven model predictions of KSS ranged between KSS 3.45 and 8.27 during the study period, whereas ratings from participants extended across the full nine point range of the KSS scale. Approximately a third (34%) of the participants' ratings were within 1 KSS point (above or below) the KSS prediction at the time the rating was made, which is highlighted on the graph as the portion of ratings between the two bold red lines. Approximately two thirds of the participants ratings were associated with a KSS prediction that was within 2 KSS points (above or below) the self-report rating, at the time the rating was made. Approximately a third of participants' ratings were associated with KSS predictions which were over 2 KSS points above or below the self-reported rating, and thus represent occasions where the model predictions are associated with a substantial margin of error. As may be seen from the upper left and bottom right quadrants of Figure 17, there was a clear asymmetry between occasions where model predictions substantially 'underestimated' participants' reported KSS levels (28%), and occasions where participants' KSS levels were overestimated (6.74%).

Self-reported involuntary sleep during flight

75 reports of involuntary sleep were submitted via the 2-Way App during the one month study period from pilots from nine different companies. Four of these reports came from pilots that had not opted into the study, and hence were excluded from further analysis. The remaining 71 reports came from long haul pilots (n=38) and short haul pilots (n=33) from nine different companies. In terms of pilot flying role, there were 30 reports from captains (including roles such as training captain) and 41 reports from first officers (including senior first officers). Forty-two reports related to the participant having involuntarily fallen asleep themselves, 29 related to the other member of flight crew having been reported to involuntarily fallen asleep. Within these, there were nine occasions where there were reports of involuntary sleep for both crew members for the same flight.

In order to calculate the rate of involuntary sleep report per flight hour, pilots' work schedules were required. Of the 71 reports of involuntary sleep collected, 20 were reported by pilots submitting incomplete roster information, and as such the per-flight hour calculations could not include these reports. The rate of involuntary sleep per flying hour is hence calculated as the number of reports (51) divided by the total flying hours (6986.67), which is 0.0073. This calculation provides a rate of 1 report of involuntary sleep on the flight deck per 137 flying hours. The rate of involuntary sleep events reported for both flight crew during the same flight was 1 report per 1818 flying hours.

Involuntary sleep data submissions were reviewed for any additional commentary included in the optional free text narrative box of the 2-way-App form. Brief comments or snippets of narrative information accompanied 27 of the involuntary sleep reports (e.g. *"We both felt very sleepy about an hour before top of descent"*; "*Captain was wordlessly nodding off during the whole return flight*"). Where phase of flight for the involuntary sleep was mentioned (n =11), seven reports referenced top of descent or during the holding procedure, three referenced the cruise portion / entirety of the flight and one during the climb period. Contributing factors cited in the 27 reports included being called out from standby, departing late (due to technical problems or aircraft changes) resulting in increased time continually awake, roster transitions between early and late duties, and accumulating fatigue from consecutive early start times. However, it should be noted that overall the data quality captured from the free text narratives was low, providing limited additional content to review.

The data were also investigated to see whether there were individual (age, chronotype) or roster-related predictors of self-reported cases involuntary sleep, using a general linear mixed model analysis for binomial data. However, the outputs indicated both non-significance of predictor variables but also more importantly singularity (as evidenced through perfect -1 correlations within random effects), indicating that the dataset did not contain sufficient cases of involuntary sleep to discriminate between predictors of involuntary sleep and predictors of 'non' involuntary sleep with individual pilot as the grouping level.

4.7 Discussion of Study 2 results

Against the context of increasing performance-based regulation of the hazard of pilot fatigue in commercial aviation, it is essential to understand the rates of both self-reported and predicted sleepiness levels occurring during flight across the entire aviation industry, using the same scientific measures. Field studies assessing the rates of sleepiness or sleep-related phenomena tied to operational flying hours in aviation are often difficult in practice to conduct with participants from a wide array of different companies and types of operation. The data in the present study provide the first benchmark description of predicted fatigue risk rates associated with British airline pilot rosters and actual occurrence rates of high levels of sleepiness and involuntary sleep during commercial flight.

Predicted sleep opportunities and analysis of duty shift timings prior to duty

A quarter of all flying duty periods were predicted to be preceded by a main sleep opportunity of less than six hours, and up to 43% predicted to be preceded by sleep opportunities of less than seven hours. There is a broad scientific and medical consensus that most healthy adults require between 7-8.5 hours' sleep per night to feel well and maintain full cognitive effectiveness (Kronholm et al., 2009; NASA, 1996). Shorter sleep durations of around six hours per night are moreover likely to cause meaningful sleepiness or impaired performance in the average shift worker (Åkerstedt & Wright, 2009; Van Dongen et al., 2003). Against this context, the biomathematical modelling of participants' rosters predicted that a substantial proportion of commercial flying duties were preceded by insufficient sleep opportunities. While individuals vary in their need for sleep and their trait vulnerability to the effects of sleep loss (Caldwell et al., 2005; van Dongen, 2006; van Dongen & Belenky, 2009) inadequate sleep prior to duty may hence be a prominent source of schedule-driven fatigue risk exposure for many pilots within UK airline work schedules. As expected, predicted cumulative sleep loss prior to duty as a schedule-driven fatigue risk exposure was almost exclusively seen in short haul operations, with 11.3% of short haul flying duties predicted to be preceded by three consecutive sleep periods of less than seven hours and 2.1% predicted to be preceded by three consecutive sleep periods less than six hours. Although a relatively much smaller proportion of our sample, consecutive periods of sleep loss present an accumulating fatigue risk for pilots from a safety perspective (Belenky et al., 2003; Van Dongen et al., 2003). Such sleep loss may be of particular concern during operations since chronically sleep-restricted individuals may be less aware of their level of fatigue-related impairment than more acute forms of sleep deprivation (Williamson et al., 2011). Whilst various regulatory principles concerning schedule design emphasize the need for adequate rest in-between duties, and at least an eight hour window of sleep for pilots between duty periods (ORO.FTL.235), the present findings suggest that in over 40% of actually flown schedules, such ‘out of work' gaps are not predicted to provide eight hour sleep opportunities in terms of their biological plausibility for the average individual.

For both long and short haul pilots, 16.2% of duty start times in this study were shifted three or more hours relative to previous duty start times. As a schedule feature, shifting shift times are likely to interfere with the length and consistency of sleep wake patterns in pilots due to both slow circadian rhythm adaptation to different waking hours, and the difficulty for pilots to adopt consistent coping strategies for abrupt shifts in their sleep patterns within their home lives. Within our sample of 1295 duty periods, duties that started or ended during the known window of circadian low periods furthermore made up over a third of all duties, with 310 (24%) having check in times between 00:00-06:00. The problem with having to get up earlier than usual is that it is very difficult, if not impossible to fall asleep sufficiently early the night before in order to compensate for the early rising time (even when the duty schedule permits), due to lack of adequate homeostasis sleep pressure. Since previous research with pilots and other shift workers has indicated that early duty start timings in particular dramatically restrict the amount of sleep obtained and increase on-duty fatigue levels (Ingre et al., 2008; Roach et al., 2012), this schedule-driven fatigue risk exposure may hence be a particularly important area for practitioners to target for reduction or provide mitigation measures for within the surrounding duties.

Predicted time of continual wakefulness at last landing

Biomathematical analyses estimated that on average short haul pilots were likely to have been awake for 11.56 hours at last landing, (*SD* = 3.76), compared with long haul pilots with an average 18.31 hours, with a large deviation around the mean (*SD* = 5.6). In terms of safety risks, these findings indicated that pilots operating long haul duties may be particularly at risk of sleepiness and fatigue-related performance decrements towards the end of their duties. Extended periods of wakefulness after about 16-18 hours of wakefulness have a profound impact on alertness levels and performance decline and hence are linked to elevated fatigue-related risks in human operators (National Research Council, 2011; van Dongen et al., 2003; Williamson et al., 2011). Extensive evidence from both road crash statistics and driving simulator studies further suggest that this elevated fatigue risk exposure at last landing may be important for pilots during their commute home, particularly where duty ends coincide with circadian lows (Horne & Reyner, 1999; Ingre et al., 2006; Reyner & Horne, 1998) . In the present study biomathematical estimations of continual hours of wakefulness did not apportion in-flight rest opportunities, an assumption that was estimated to be relevant to 5% of the flying duties analysed in this study. However, in terms of approximating fatigue-related exposures from work schedules, it was felt that a greater source of inaccuracy would stem from modelling in-flight sleep opportunities where the timing, availability and utility of such opportunities was not known. Future investigations would benefit from the collection of precise in-flight rest data in terms of both possible sleep opportunities and whether such opportunities resulted in sleep across a variety of longer operations. In addition, there are rules and principles governing pilots' hours of work so that they avoid 18 hours of continuous wakefulness during their duties (Civil Aviation Authority, 2016: GM1 CS FTL.1.225(b)(2)). However, our modelling analyses (not taking into account any diversity of sleep-wake strategies prior to duty), still estimated that continual wakefulness associated with pilot work schedules may be a prominent fatigue risk exposure in actually flown schedules, particularly in long haul operations. Moreover, continual hours of wakefulness at the time of participant rating was also found to be a significant predictor of reported KSS (discussed further below).

KSS predictions and self-report ratings during Flying Hours

Whilst the average predicted sleepiness scores at key operational points (duty start, last landing and duty end) were overall not indicative of severe KSS sleepiness scores (KSS >7), a substantial proportion of flying hours were associated with predicted and reported sleepiness at levels that may represent a risk to flight safety. For research informing the scale and severity of fatigue hazards experienced by pilots during flight, it is important to underscore what such KSS levels mean. At KSS levels above 7, laboratory research has shown that there is a marked increase in EEG-defined sleep intrusions and long eye lid closures, suggesting that at this level, it is difficult - even with high motivation levels - to stave off sleep (Åkerstedt et al., 2014; Anund et al., 2009; Ingre et al., 2006; Reyner & Horne, 1998). KSS levels of 8 and above are of particular concern since this level on the alertness-sleepiness continuum is associated with markedly increased microsleeping or involuntary sleep intrusion risk, escalating performance decline and increased collision risk in other domains (Åkerstedt et al., 2014). Against this context, the present study findings that pilots submitted one in-flight rating ≥KSS 8 per 59 flying hours indicated that very high levels of sleepiness in flight occur, and likely occur far more routinely than had previously been documented. When compared against the biomathematical model predictions, the present study findings suggest that 10% of flying hours were associated with KSS predictions of 7 or above, within which 3% were associated with predicted fatigue levels of KSS 8 or above. Reported and predicted KSS per-flight hour rates are not directly comparable, since KSS predictions per flying hour can be derived from continuous model predictions, whilst the reported KSS rates of elevated sleepiness levels per flying hour represent discrete occasions where the participants both felt subjectively sleepy *and* submitted a rating. Hence, unless comparable intervals of ratings and predictions are mandated in the methodology, per-flight hour reported rates are methodologically likely to be a subset of the predicted KSS per-flight hour rates, because opportunities to complete a rating are not always available, convenient or safe. As such, had more frequent submissions been received, the frequency of KSS ratings ≥ 8 may have been greater than the 1 per 59 flying hours observed. Notwithstanding this qualification, it is worth noting that both predicted and reported KSS levels of elevated sleepiness (1 in 10 flight hours predicted ≥ KSS 7, verses 1 rating per 33.1 flight hours ≥ KSS 7) are of a similar magnitude, and appear high from a safety point of view, given the neurobehavioural deficits evidenced through laboratory work at these levels of sleepiness.

Comparisons of fatigue risks with parallel medical standards for the human component in aviation

Within the commercial aviation industry there has not been universal agreement at what rate of occurrence reported or predicted fatigue risks such as these become unacceptable from a broader safety point of view, for a number of reasons. First, there is limited experimental evidence investigating sleep loss, both chronic and acute, and its impact on multi-crew commercial flight performance. Due to advances in the high reliability standards of aircraft automated systems that help control the trajectory of flight for most of the cruise portion, further research is certainly needed to better understand the relationships both between increasing pilot sleepiness levels and unsafe individual pilot performance, and the resultant impact on the overall safety of flight. Acceptable levels of schedule-driven fatigue risk will furthermore inevitably sit in conflict with commercial productivity and optimisation of crew. As such, the risk appetites of different industry stakeholders regarding acceptability of elevated KSS predictions during flight can, and often do differ within both prescriptive and performance-based regulatory limits (European Aviation Safety Agency, 2014). However, where parallel standards of the risk of incapacitation of flight crew do exist is in the medical incapacitation rate, which is set at a target rate of less than one occurrence per 1,000,000 hours. The study findings indicate that 10% of flight hours were predicted to be associated with KSS 7 and above, and 3% of flight hours predicted to be associated with KSS 8 and above. This comparison highlights a fundamental disparity between the acceptable probabilities of schedule-driven on-duty fatigue risk, and medical incapacitation risks, despite both risk rates relating to significant in-flight functional impairment of crew. In terms of reported rates of severe levels of sleepiness, the present study findings of 1 rating ≥ KSS 8 per 59 flight hours also appear high against this standard, even though sleepiness may be somewhat more 'reversible' by sleep, if this is possible during flight, than many types of medical incapacitation. Indeed, sleep-driven 'microsleeping' or involuntary sleep attack events are both the most direct consequence of physiological sleepiness, and also referred to as categories of in-flight medical impairments (Dejohn et al., 2004), since the functional impairment to pilot performance is likely to be significant at very high levels of sleepiness (e.g. degradation of visual awareness, attentional lapses and sleep attacks). Within this study there were 71 reports of involuntary sleep during the one month period, and with the data available to calculate a per-flying hour rate, this equated to 1 report of involuntary sleep on the flight deck per 137 flight hours. Compared with the target medical incapacitation regulation standard of no more than 1 occurrence per 1,000,000 hours, the magnitude of this approximate 7,000 fold difference would appear a non-trivial difference in occurrence rate. On four of these occasions involuntary sleep was reported for both flight crew during the same flight, equating to a rate of possible overlapping involuntary sleep events and simultaneous sleepiness in both flight crew of 1 per 1818 flight hours. The recording of such events are reliant on self-report, and as such it is likely this rate reflects an underestimate of actual occurrences as previous research using sensitive objective recording methods of sleepiness such as EEG in both flight crew has revealed that there may be a number of occasions where pilots involuntarily fall asleep and reawaken without knowledge of the event (Wright & McGown, 2001, 2004). Clearly, simultaneous sleepiness within both flight crew is of elevated safety concern since, from a safety systems point of view, as it represents a form of common mode failure of a safety critical system (Downer, 2009). This means that the assurance otherwise provided by having independent risks of failure from multiple crew may be compromised where sleepiness risks are similar for pilots flying the same work schedules. Operator fatigue may therefore constitute a more insidious and common source of pilot impairment (Caldwell, 2005; Eriksen & Åkerstedt, 2006; Gander & Signal, 2008; Petrilli et al., 2006), more dangerous than an obvious, complete incapacitation where impaired and potentially unsafe performance goes undetected for an extended period. It should be mentioned that there is has been some debate as to whether the medical incapacitation rule is over- conservative (Mitchell & Evans, 2004). However, the current study findings still highlight a substantial disparity between current acceptable risk standards with respect to pilot functioning and impairment during flight. As some regulatory bodies have pointed out, knowledge of what medical conditions or in-flight impairments are affecting pilots and possibly contributing to a crash or incident '…would be useful in assisting the on-going evolution of the aeromedical regulatory process' (Australian Transport Safety Bureau, 2007). As such, it seems important that continued research into pilot fatigue and sleepiness events during actual flying hours are monitored against similar risk standards to other forms of in-flight impairment and incapacitations.

Comparison of predicted and reported KSS

Overall, there was significant relationship between predicted and reported KSS scores. Comparison of the SWP model predictions and KSS ratings revealed that as predicted KSS scores increase by 1, reported KSS increase by 0.55 - 0.6, and this positive relationship in KSS response patterns was broadly similar across pilots. This was a statistically significant relationship and hence shows there is value in using the SWP model for predicting reported KSS. However, although the prediction relationship did not appear to differ much between pilots, there was a substantial proportion of variance in pilot ratings not accounted for by the model.

Exploratory analysis on potential roster-related variables that may add to the 2-process model accuracy found that both consecutive early and late duties within the prior 72 hours were significant predictors. This finding suggests that there may be a cumulative impact of sleep loss arising from consecutive duties that is not accounted for by the 2 process model, and accordingly wake-time KSS predictions may tend to decrease in accuracy across a block of consecutive back-to-back early or late duties. Put another way, on their own, KSS predictions for early or late duties on day 1 and 2 of a block may be more accurate than on days 3, 4 and 5. This finding supports previous research on bio-mathematical fatigue models that suggests that in general they are less accurate for predicting the effects of chronic sleep restriction (Dawson et al., 2011; Van Dongen et al., 2003) and do not always take into account sufficient prior sleep/wake history in formulating sleep and wake predictions (Mallis et al., 2004). Indeed it seems that whilst recovery from total sleep deprivation may be predicted with reasonable accuracy, recovery following chronically restricted sleep is typically much slower in real life workers than would usually be predicted by bio-mathematical models of fatigue (Dinges, 2004; Ramakrishnan et al., 2016). This finding also provides some support to other validation applied field research findings that both time-of-day and days-on-shift influence the rate at which fatigue accumulates (Riethmeister et al., 2019; Roach et al., 2012; Spencer & Robertson, 2002). As such, there may be additional value for bio-mathematical fatigue model vendors using the 2-process model to investigate changes in KSS rating profiles over consecutive early or late duties, in order to see if there are accuracy improvements to be made for KSS predictions that fall towards the end of a roster block.

Two other roster-related variables investigated were sleep hours within the last 24 hours, and continual hours of wakefulness. Schedule types such as long haul duties and some short haul duties that have late duty start times are likely to result in extended hours of continual wakefulness at the time of rating. A priori, it was assumed that neither of the variables would significantly add to the existing KSS predictions, since both should already be accounted for by the 2-process model. Interestingly, as shown in *Models 5* and *6*, continual hours of wakefulness at the time of rating was a significant predictor of reported KSS, in addition to the variance already modelled by the 2-process model itself. A possible explanation of this finding is that the accumulation rates of sleepiness may differ depending on the tasks involved and the context of the activities undertaken, which is not accounted for in the underlying 2-process model. For example, there might be an additional fatiguing effect of being awake and doing piloting tasks (during both complex, cognitively demanding periods and monotonous periods) that acts to increase the levels of sleepiness experienced by pilots over longer periods of wakefulness. Two sources of evidence provide some support to the notion that certain tasks and contexts may induce additional sleepiness. The first and strongest evidential support is provided by the fact that most real road and driver simulator studies indicate a time-on-task effect of KSS, that occurs irrespective of time of day, but is more pronounced during night driving (Åkerstedt et al., 2013, 2014; Anund et al., 2009; Sagaspe et al., 2008; Sandberg et al., 2011). Secondly, there is also some indirect evidence in support of context-dependent differences in KSS ratings across waking hours through comparisons of normative data using the KSS in different settings (Åkerstedt et al., 2017). Although such studies have tended to be rather small, they have suggested that increased levels of KSS have been reported during wake hours of a working day in contrast to a day off (Åkerstedt et al., 2017; Söderström et al., 2006). Further evidence of task influences include ‘boring’ low stimulus environments which also increase subjective sleepiness, as evidenced through long stretches of train driving through non built up areas (Ingre et al., 2004) and the temporary rebounding of alertness that is seen when breaks are taken from monitoring of computer screens, and subsequent return to pre-break sleepiness levels following resumption of the monitoring task (Gillberg et al., 2003). As such, this result of additional predictive power arising from continual hours of wakefulness at the time of rating may be due to the influence of time on task, the cognitive load of tasks and potentially a number of different context-specific features specific to the flight deck environment, over and above the expected rise in homeostatic sleepiness for the average person that would be predicted from the 2-process model. Taken together, the significant improvement of model fit to the reported KSS data found for Model 6 suggests that additional investigation efforts into roster-related variables, particularly those that may be simply determined from the roster itself (such as the impact of prior consecutive early or late duties) is worth exploring in future validation work on the 2-process model equations.

Finally, the data were examined to ascertain whether model estimates were more likely to underestimate or overestimate reported KSS. Comparisons of the predicted KSS at the time of rating against the reported KSS ratings submitted at that time, showed that the 2-process model was substantially more likely to underestimate the reported KSS submission. This is perhaps theoretically understandable, since the 2-process model used to generate the KSS predictions only takes into account sleep and wake parameters, rather than any other factors which may potentially impact an operator’s sleepiness levels. This implies that when predicted KSS for a particular roster is shown to be high (KSS 7-9), there is a strong likelihood that actual KSS experienced will also be the same or higher. Given the incremental risks associated with operating at high levels of fatigue, this hence represents an important observation for fatigue risk prediction and how biomathematical model outputs are understood by practitioners and applied to organisational fatigue risk policies.

4.8 Strengths and limitations

The main strength of the present study is that the results are based on a large dataset collected from the pilots of eighteen UK airlines under naturalistic working conditions. The results are hence likely to reflect the roster-related fatigue risk exposures and pilot sleepiness rates in the UK more far more reliably than field studies conducted with one or two companies or general fatigue surveys among pilots which tend to lack precise timescales and validated measures. Indeed, a key benefit of the methodology of Study 2 was the fact that KSS scores denoting the pilots’ experience of the severity of sleepiness were time-stamped and could be meaningfully compared against biomathematical modelling predictions of the pilots’ rosters using the same measure and appraised with respect to the sleep scientific literature. This strength is particularly notable, since to date, there is still no consistent fatigue risk measurement that has been adopted across the UK airline industry, and as such airlines tend to vary in how they appraise and report fatigue risks and events (European Aviation Safety Agency, 2014). This means that at the industry level, there has been at best, a fragmented public database with respect to fatigue risk exposure data, and there is likely to be few common measures for the Civil Aviation Authority to compare between airlines beyond ‘number of fatigue reports’ submitted via the MOR system. The analyses enabled comparisons between the flown work schedules from long and short haul operations and the experience of pilots from different airlines, which together provided insights into the sleepiness risk rates across a sample broadly representative of UK commercial aviation industry, and not just specific operational routes. Since the large-scale collection of individual sleep habits is often not feasible across an entire workforce, the use of biomathematical modelling to provide not just on-duty alertness estimations, but also aggregated summaries of predicted sleep and continual wakefulness opportunities prior to flying duties is furthermore a practical data category of fatigue risk exposure that can be determined from pilot work schedules. This study findings highlighting a considerable proportion of reduced pre-duty sleep opportunities associated with pilot rosters hence underscore the value of such data insights in informing how well different regulatory rules seeking to control such risks actually perform in practice.

Based on a review of comparable aviation fatigue literature, the present study is the first of its kind to combine fatigue analysis of flown rosters via biomathematical modelling and collection of self-report data using the same measures in order to produce per flight hour estimates across pilots from 18 different UK airlines. This study is also thought to be the first of its kind to combine the collection of predicted and reported KSS and sleep events from pilots over a month long study period. Indeed, even amongst other occupational settings, one of the key vulnerabilities of many applied fatigue research initiatives is that the timescales for the study period are restricted to two weeks or less. This means that the data available to assess particular roster combinations such as consecutive early or late duties will be more restricted where the timeframes are more restricted. This study also demonstrated that voluntary data submissions can be maintained over a month long period, particularly where submissions are quick to complete and benefit from easy to use technology. As such, this finding may have broader methodological relevance to future research initiatives in applied fatigue research in different occupational environments.

There are several limitations of the present study. Various compromises were made in the design of the methodology to try to increase the potential sample size and maintain levels of participant engagement during the month long study period. Such compromises were informed by the stakeholder consultation process at the outset of this project and principally acted to minimize the amount of information required from participants. First, beyond the submission of ratings via the 2-Way app and transfer of roster data at the end of the study period, there was limited additional information required from pilots. This meant that additional health (e.g. sleep, physical health factors, mental health conditions), lifestyle (e.g. smoking), or further relevant workload factors were not collected but may have positively or negatively influenced sleepiness levels, cases of involuntary sleep, or affected the relationship between predicted and reported KSS. Indeed, sleep disorders, smoking and increased workload have all been found to negatively affect sleep and fatigue levels (Magnavita & Garbarino, 2017; Mcnamara et al., 2014; Wong et al., 2008), whilst stress, depending on which aspects are measured, appears to have differential effects on sleepiness (Åkerstedt et al., 2014; Axelsson et al., 2004; Ekstedt et al., 2009). Thus individual differences in sleepiness profiles related to such factors might exist in the pilot population and should be considered as important to investigate in further studies. A further compromise was to instruct participants to provide six or more KSS ratings spread over their hours of wakefulness during the study period. This meant that there was no systematic collection of ratings at particular operational points of significance such as ‘top of descent’ which have been used in other aviation fatigue monitoring studies (Gander et al., 2014; Sallinen et al., 2017, 2020) and which are considered operationally relevant (as they are thought to denote risk-based changes in safety with respect to the aircraft’s proximity to the ground and complexity of task). The flexible rating guidance may have also influenced the large variation in total submission rates across the pilot sample, which saw a small number of pilots (n = 13) submit less than ten ratings across the study period, with significantly higher scores compared with participants with greater submission rates. As such, the flexibility of the methodological approach may be considered a limitation of the study method, although this may be balanced against the possibility that rating flexibility may have incentivised more pilots overall to participate in the study or continue submitting data across the study period.

Although higher sample numbers are always desirable, the large sample of pilots who opted in were found to be representative of the pilot population as reflected by the BALPA membership numbers. On the other hand, the pilot sample was not based on randomization and non-BALPA members would have likely been missed from the communications highlighting the ability for UK pilots to participate in the study. Although BALPA membership among UK airline pilots was estimated to be around 85-90% of the UK airline population and thus represent the vast majority of commercial airline pilots, both these factors should be recognised as potential limitations to the representativeness of the study. In order to feasibly sample a larger cohort of pilots and duty patterns, biomathematical model predictions were also relied upon to provide the estimates of sleep length and timing, based on the pilots' working schedules as inputs. Biomathematical models estimate the length and timing of sleep where the biological drives permit based on the duty hours alone, which means that if sleep is predicted to be biologically probable and there is no work duty, it will be apportioned. Hence, in this regard, predictions of sleep length will not take into account other work not detailed in the work schedule, additional travel or out-of-work hassle factors that may extend the individual's hours of wakefulness further. Furthermore, estimations inevitably cannot reflect the diversity of sleeping patterns that pilots may achieve. Whilst additional individual data collection via objective recording instruments such as actigraphy would hence have been desirable to check the accuracy of predicted sleep lengths to individual sleep experience, this was simply not feasible for this large-scale exercise. These data were also not considered imperative for this study, since a key aim was to investigate the average likely sleep opportunities that working schedules are likely to provide from the SWP (with underlying two process model). Continuous objective measures of sleepiness to complement pilot ratings during flight would have also been desirable since KSS ratings share the same limitations of all self-report measures (such as vulnerability to conscious or unconscious bias in response patterns). However, such measures were also not considered practical for large scale exercises such as the present study. Analyses of the study data showed that participants were, in accordance with instructions, submitting ratings across the KSS scale both during flying hours and non-flying hours and submission rates were sustained across the study period, suggesting that the self-report data were not obviously jeopardised by any clear floor or ceiling effects. Moreover, there was no indication of odd rating patterns or large numbers of repeat submissions of involuntary sleep from individual pilots. However, where feasible, additional measures to monitor pilot sleepiness experience should be a methodological consideration for future studies, particularly if both pilots could be monitored for indications of simultaneous physiological sleepiness and the time course of their development. A further limitation of this study method was that the manual input of rosters into the SWP was enormously resource intensive, which may have an impact on the ability for other researchers to replicate this study across such a wide array of airlines. Since roster types provided by pilots varied by airline and also by format in terms of the data transferred across, the data input stage was considerable in terms of the research time involved. Thus future research that can develop automated analysis of raw roster data would be highly desirable, although it may not surmount the broader problem of inconsistencies between different organisations in terms of how input variables are coded and displayed. It should be recognized that other commercial biomathematical model vendors may provide some automatic 'bulk analysis' function for particular roster formats. However, it is not known whether these commercial models facilitate automatic analysis of all types of roster and whether prediction data could be examined for public research purposes in the same way as the SWP model providers permit. A further consideration for this study is that it was conducted during the August month period, which is likely to be part of the busy season for many airlines. Hence the work schedules examined may reflect an increased duty workload for some of the airlines than other seasons within the year. Follow up studies investigating the same variables would be useful to indicate if there is substantial seasonal variation across long haul and short haul operations schedule-driven fatigue risks and experienced levels of sleepiness.

The main limitation of the study results related to missing data from pilots who had opted into the study (in terms of roster information and/or demographic details provided). Of the initial opt in sample of 294 pilots, 184 pilots provided ratings, but only 95 provided full roster and rating information, and as such there was a reduction in the data available to inform per-flight hour estimates. For field studies, it is to be expected that some participant attrition occurs that may impact the collection of data, but in this case it is possible that the retrieval and sending of rosters may have been either been more challenging for some pilots (due to particular features of their company's systems) or may have simply been forgotten at the end of the month long period. Future research efforts may hence benefit from exploring different methods for roster data capture or indeed coordinating with individual safety departments to receive roster information from the organisation itself rather than relying on participants to transfer these data. The dataset furthermore did not contain sufficient cases of involuntary sleep to discriminate between predictors of involuntary sleep and predictors of ‘non’ involuntary sleep with individual pilot as the grouping level. This finding suggests that further research with larger datasets collecting incidence of involuntary sleep is needed to provide sufficient cases to statistically assess.

In conclusion, the biomathematical model analyses in the present study indicated that a substantial proportion of flown pilot working schedules are likely to be associated with insufficient sleep opportunities prior to flying duties. High levels of in-flight sleepiness were both reported by pilots and predicted by biomathematical model estimations using the pilots’ work schedule times as inputs. This study provides a benchmark of these rates against the Karolinska Sleepiness Scale. In order to provide context to both predicted fatigue risk rates and self-report rates of high levels of sleepiness, it is suggested that attempts should be made to find common risk denominators, such as per-flight hour rates of occurrence, related to the way that in-flight incapacitations and impairments of pilots are appraised in commercial aviation. Where this comparison is made, the present study findings suggest there is a non-trivial difference of risk tolerance between an existing target medical incapacitation rate for pilots and both the predicted fatigue risk rates associated with pilots’ schedules and pilots’ reported rates of elevated sleepiness and involuntary sleep during flying hours. Since sleepiness risks during flight are likely to not be independent between flight crew undertaking the same duties, this finding further suggests substantial efforts to reduce fatigue risks during normal flying operations may be required in order to meet the existing target safety standards of the human component.

Chapter 5. General Discussion

5.1 Research approach

Operator fatigue is widely recognised as an intrinsic human factors hazard affecting work environments which rely on shift-work and extended work hours. Across the last five decades, a wealth of studies from a range of transport and occupational settings have indicated that sleep and circadian factors are related not only to changes in cortical function and performance impairment, but also to increased injury and accident risk (Williamson et al., 2011). However, whilst the effects of increasing levels of sleep deprivation on neurobehavioural performance have become relatively well established, links between operator sleep-driven fatigue and safety outcomes are not as straightforward to summarize and many gaps exist in the safety and sleep science applied field literatures. This is because the safety impact of impairments to human performance (via error, incapacitation or otherwise) is contingent on the operational environment, the safety-criticality of the human component, and the reliability of the safety buffers or mitigations that exist to mitigate any form of human failure. Existing real-world data on the impact of operator fatigue to operational safety furthermore depends to a great extent on reporting culture and whether reported incidents or accident investigations include a routine assessment of sleep-wake history or other fatigue risk indicators.

The relationships between fatigue and safety outcomes have been perhaps best understood in the road transport sector, aided by systematic research into sleep and circadian factors in crash statistics, driver sleep-wake histories and sleep-deprivation simulator studies. Such work has helped provide important ‘rules of thumb’ of risk thresholds around driver fatigue by clarifying associations between sleep deprivation levels and increased subjective sleepiness, physiological changes (including microsleeping risk), driving performance decline and accident risk. By contrast, in environments which are extensively safety-netted and highly regulated, the relationships between operator fatigue, breaches in sub-system safety and whole-system accident risk have been arguably more complex to investigate. ‘Ultra-safe’ industries such as scheduled commercial aviation and nuclear power are supported by high reliability systems with multiple layers of operational defences which collectively act to reduce the probability of having an accident attributable to a unique cause (Gander et al., 2011). In these industries, accident rates are orders of magnitude lower (e.g. <1 accident per one million events[4]) than other lower-reliability regulated systems such as commercial

---

[4] Amalberti’s (2001) use of the term ‘events’ in this context relates to industry specific units (e.g.number of flights, operations, running time). These units vary according to industry or transportation mode, garnered from statistical databases from different industries.

driving and chemical industries (±1 accident per 1000 - 100,000 events) (Amalberti, 2001). Hence, on a daily basis, the impact of pilot fatigue (or indeed any cause of human performance detriment) to overall system safety is likely to be greatly reduced compared with less reliable occupational environments. However, safety concerns regarding pilot fatigue still exist even in the context of modern commercial aviation operations, despite quantitative restrictions on pilot flying hours and a regulatory framework aimed at controlling excessive fatigue risks to flight crew. This is because despite their low probability, aviation accidents have the potential to cause catastrophic loss of life, alongside enormous societal and financial costs. Furthermore, as described in Chapter 1, serious aviation incidents still occur and continue to highlight weaknesses in regulatory and organisational fatigue risk management defences. As recently as 2017, a serious near-miss occurred when an Air Canada plane was thirteen feet from crashing into four passenger occupied planes waiting on a taxi-way. In this serious incident, the investigators not only identified fatigue as a major contributory factor affecting the captain, but also highlighted the legitimate system safety risks of 1) extended (but permissible) waking hours for pilots on standby-duties and 2) fatigue-related impairment in not just one, but both flying pilots (NTSB, 2017).

This context suggests that research on fatigue in commercial aviation and other ultra-safe environments must seek to investigate signals of elevated fatigue risk at multiple organisational defence layers (both distal and proximal), to understand when and where assumed safety thresholds may be threatened and provide new data on how frequently such thresholds are breached. This is particularly important for the regulatory context in aviation which increasingly seeks to move from traditional hours of service controls to performance-based fatigue-risk management. This regulatory approach hence relies heavily on gathering and monitoring operational fatigue data from front line staff in order to appraise the effectiveness of the derogation of risk assessments to organisations to undertake and manage themselves. Such data may further be crucial for the assessment of any changes in regulation that may fundamentally alter the types of working pattern pilots routinely undertake (as in the case of changes in flight time limitation rules and derogations from rule sets during certain periods of operational uncertainty, such as during and following the covid pandemic (IATA, 2020). The approach adopted in this thesis has been to use the theoretical Fatigue Risk Trajectory framework to help dissect the practitioner-led question of ‘how tired is too tired for safe flight?’ and structure two research studies that aimed to provide new data on real world fatigue risk exposure and sleep incidents across normal UK commercial aviation operations, and greater clarity on the characteristics of existing fatigue-related commercial aviation safety occurrences in the UK. Structured this way, the links between fatigue risks and safety outcomes can be understood more readily in the context of ultra-safe environments and manifestations of fatigue risk can be interpreted with respect to evidence from sleep science on neurobehavioural fatigue-related decline and existing safety thresholds for the human component in commercial aviation.

5.2 Evaluation of study objectives

To appraise existing real-world safety data on fatigue risks in aviation using the FRT framework as guidance, Study 1 sought to produce a review of fatigue risk data derived from mandatory occurrence safety reports (MORs), the main safety occurrence data collected by the Civil Aviation Authority within UK commercial aviation. This aim of this study was to ascertain what research insights could be drawn from current existing safety report data on fatigue risks in commercial aviation. The objectives were threefold;
1) To assess the quality and coverage of fatigue risk data (FRT levels 1-4) captured by the MOR database;
2) To investigate whether there were discernible associations between fatigue risk parameters and increased risks to pilot performance and emergent system safety
3) To evaluate any emergent insights, relationships, common themes or characteristics that may be gained from an examination of fatigue-related MORs.

As discussed in Chapter 3, Study 1 provided an assessment of fatigue risk data coverage within MOR reports using the framework of the FRT as guidance. For this study, part of the preparatory work involved the development of specific data metrics against the FRT framework drawing on the applied sleep science literature. The utility of the coding structure developed was that MOR report narratives could be thoroughly reviewed for the inclusion of different categories of fatigue risk information, but also distinctions of data quality could be made with respect to whether fatigue risk data, where included, would be sufficient or precise enough to enable post-hoc bio-mathematical analysis or comparisons between reports across the entire dataset. This approach allowed the first objective to be met yet the analysis suggested that overall, the quality and coverage of fatigue risk data (FRT levels 1-3) was low and inconsistent across reports. Against the framework of FRT, overall the MOR reports did not include sufficient data to allow for distal and proximal fatigue risks surrounding the safety occurrence to be quantitatively investigated (such as precise work pattern timing information, sleep-wake history of the reporter or time of day of the occurrence). As such, objective 2 was unfortunately not pursued as the ability to quantitatively assess the relationships between fatigue-related risks and adverse safety occurrences was not deemed possible.

Although limitations in the MOR data meant that objective 2 could not be explored within Study 1, there are nevertheless clear methodological benefits to post-hoc quantitative assessment of sleep and circadian parameters in occurrence data that have been demonstrated in other transport environments. For example, greater scientific understanding of the conditions under which driver fatigue may affect safety has been enormously benefitted by the assessment of crash databases. Such research has helped to identify not only clear time of day effects for sleep-related vehicle accidents (peaks around 0200–0600 and 1400–1600) (Horne & Reyner, 1995a; Horne & Reyner, 1999) but also dose-response relationships between drivers' sleep in the past 24 hours and their risk of causing a motor vehicle crash, highlighting a fifteen fold increased crash risk for those had slept less than four hours compared with drivers who reported 7-9 hours sleep (Tefft, 2018). For such studies, precise fatigue risk data is necessary for the identification of inflection points of sleep-related adverse safety outcomes for particular operational environments. Indeed, without systematic collection of data during normal operations, it is difficult to quantitatively assess how fatigue risk exposures across the commercial aviation industry may contribute to safety outcomes or incidents. The development of such evidence bases is hence highly desirable, and would provide a stronger basis for any future evaluation of the performance of fatigue risk management systems within individual companies or evaluation of the impact of flight time limitation rule changes.

Despite the challenges associated with a disparate database, there were several findings relating to objective 3. These included the convergence of three main themes highlighting common fatigue risks at the FRT levels 1-2, and new data insights on fatigue-related errors and the reported breaches to safety thresholds at FRT levels 3-4. The finding of a high proportion of fatigue-related MORs without visible operational consequences raises important questions for the retrieval of data within future MOR submissions. Whilst it can be useful to collect data on in-flight sleep-related events and symptoms since such data (where standardised scales are used) denote a probabilistic impact on pilot performance, it is argued that the MOR database should collect further information on whether the other flying crew member was affected by fatigue and why, in the reporter's view a more serious safety event did not occur as a result of their fatigue state. Future reporting requirements for mandatory occurrence reports (discussed further in the following section) are therefore suggested to yield both the specificity of data needed for fatigue-risk analyses and also provide important safety data pertaining to the impact of fatigue-related performance on flight safety.

Study 2 sought to assess the predicted fatigue risk exposure of routine UK airline commercial pilot work schedules (FRT 1) and the incidence and severity of self reported sleepiness (FRT 3) through a large scale field study of normal UK airline operations. The objectives were;
1) to provide new evidence on the aggregate fatigue risk exposure across the commercial aviation industry using bio-mathematical modelling,
2) to collect per-flight hour estimates of the incidence and severity of sleepiness experience by pilots across 18 UK airlines and
3) to trial a newly developed app to enable flight crew from different companies to self-report sleepiness symptoms and take part in research whilst at work.

As discussed in Chapter 4, the field study conducted over a one month period yielded important new data on the predicted fatigue risk exposure and reported rates of sleepiness and involuntary sleep events from UK airline pilots, which are important metrics for the assessment of the fatigue-related safety risks across the UK aviation industry. Whilst there have been previous studies which have sought to identify rates of severe sleepiness or ' pilot fatigue' experienced in commercial aviation (e.g. Houston et al., 2012; Petrie et al., 2004), it has not been possible to determine comparable rates from the findings of such studies due to the different definitions of fatigue used, and the absence of similar timescales or population groups involved with respect to operational hours. Prior research involving the monitoring of pilot sleepiness levels during actual operations has also typically focussed on addressing a very specific set of operational issues for a particular airline or type of operation (e.g. Gander et al., 2013; Gander et al., 2013; Samel et al., 1997; Wright & McGown, 2001). As such, the number and range of participating pilots from different airlines have understandably, tended to be somewhat restricted. Study 2 hence sought to address part of this gap in the literature and to assess in practice, whether a larger scale study of sleepiness during flight could be conducted with pilots from eighteen airlines across the UK commercial aviation industry.

Against the FRT framework, the collection and analysis of roster data from pilots allowed for new insights on the risk exposure arising from work hours and predicted sleepiness levels during flying hours (FRT 1), meeting the first objective. This prediction data was complemented in parallel by the collection of self-reported sleepiness (FRT 3) providing the opportunity for comparison between predicted and experienced rates of elevated fatigue risk exposure using the same sleepiness metric (KSS), meeting objective 2. As summarized in Figure 18 below, the utility of this approach is far greater clarity for fatigue risk exposure rates across UK aviation flight operations at different levels of the FRT, using the common ‘flight hour’ metric adopted in commercial aviation to provide relative rates of occurrence.

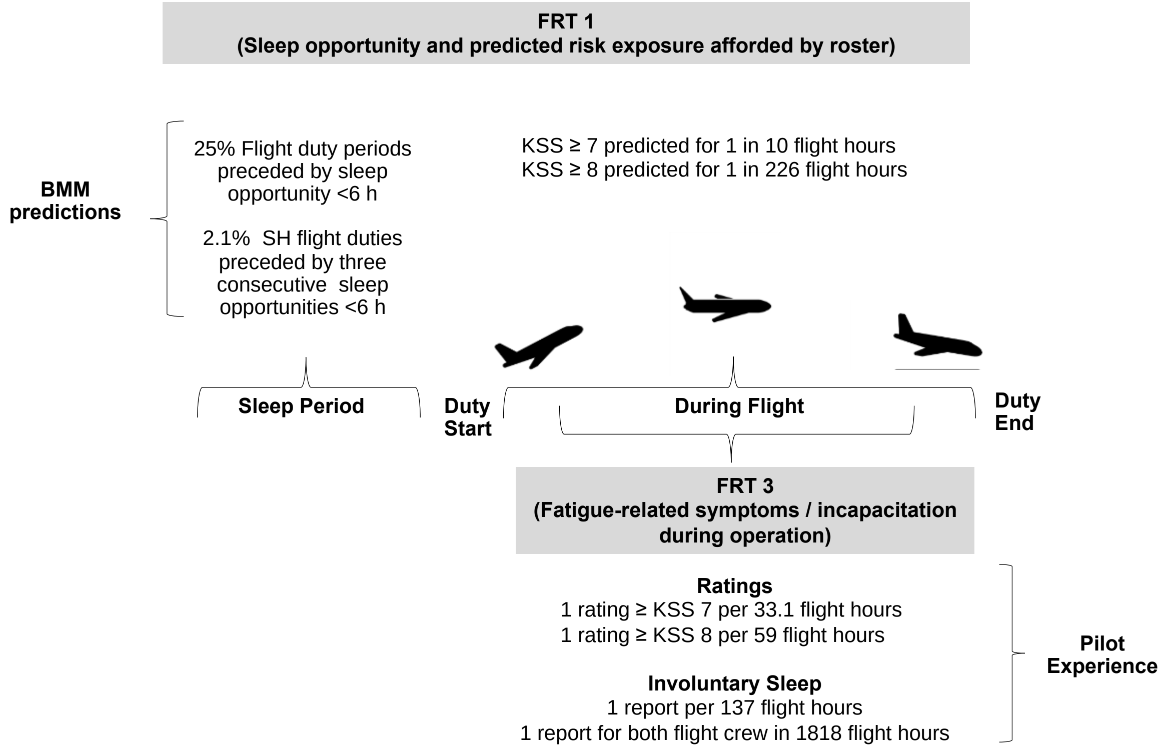


Figure 18. Study 2 headline findings of predicted and reported fatigue risk exposure rates

As discussed in section 4.7, the common flight hour metric investigated for this study furthermore allowed pilot sleepiness risk rates to be considered against the risk of incapacitation of flight crew from other human factors. Specifically a new set of comparisons can now be made for pilot fatigue rates (predicted and reported) against the medical incapacitation target rate for licensed commercial pilots which is set at less than one occurrence per 1,000,000 hours. This exercise for the first time sets fatigue risk rates in the context of other tolerated risk rates to human performance decline, and offers new data to inform the broader question of ‘how tired is too tired’ for safe flight within the context of existing human factors safety thresholds.

To enable this research, Study 2 required the development and use of an app to enable pilots to make time-stamped ratings of their sleepiness levels and reports of involuntary sleep remotely from the flight deck and whilst conducting their duties. This third objective was met through the development and successful use of the 2-Way App. These specific data fields sought to address the need for pilots to not only report sleepiness levels using the scientifically validated KSS scale but also collect the timing of ratings for subsequent comparison with their roster data. It further allowed for pilots to submit reports for in-flight involuntary sleep events. For this PhD, specific data fields of this app, guidance for use and collation and analysis of output data from the app were developed as part of the research process. As a quick and easy method of reporting, it is likely the app was imperative to high levels of engagement, as during the one-month research period, the app received over 8000 ratings from pilots from 18 different UK companies, with pilots submitting at all levels along the KSS alertness/sleepiness continuum and across the 24-hour cycle. Such information, when coupled with the participants’ rosters enabled the identification of when the alertness level occurred, and as such provide much greater precision in data capture of symptoms than traditional forms of post-flight fatigue reports.

5.3 Implications for policy and practice

In this section, the findings from Studies 1 and 2 are considered for their implications for current and future data collection and fatigue risk monitoring activities during normal commercial aviation operations and in safety occurrence reporting. The following principal areas addressed in turn are; 1) underreporting of fatigue within aviation; 2) proactive monitoring of alertness during duty 3) the use of bio-mathematical modelling for assessment of fatigue risk regulatory principles in practice and 4) the appraisal of fatigue risk data within the broader context of safety regulation within aviation.

5.3.1 Fatigue reporting

Over the last decade, across many UK and EU airlines, qualitative evidence from surveys and confidential reports with front line staff has indicated widespread underreporting of fatigue by pilots (ComRes, 2013; European Cockpit Association, 2012). Amendments in European regulations on flight time limitations since 2016 (EASA Subpart FTL) have led to a change in fatigue risk management in commercial regulation towards a safety management systems (performance-based) approach. Within this approach, pilot fatigue reports constitute important operational data for informing on the burden of fatigue within an airline (Regulation (EU) No 376/2014), and represent a significant data source for appraising the effectiveness of the hybrid of performance-based and hours-of-service controls on pilot fatigue. Coinciding with these regulatory changes, results from Study 1 highlighted a considerable climb of over 600-fold increase in fatigue-related MORs submitted to the CAA between 2015 and 2016, with a relative increase of 0.01-3.22% of all submitted MORs, which is likely to relate to either or both of these factors (increased reporting emphasis or changes to the flight time regulations).

However, a comparison of datasets from Studies 1 and 2 provides new quantitative evidence of the potential scale of underreporting of fatigue and sleepiness on the flight deck that still exists in the UK. Indeed, during the one month period of August 2017, out of the broader potential reporting population of airline transport pilot license holders (n =13847, CAA; 2017) there was one fatigue-related MOR captured by the system which cited involuntary sleep during flight (Study 1 MOR full dataset). By contrast, during the same one month period from Study 2 with an overall sample of 294 UK airline transport pilots, there were 71 reports of in-flight involuntary sleep submitted through the 2-Way App. To provide a crude estimation of this difference in reporting rates over fatigue-related in-flight incapacitation, using the assumption that each license holder in the population flew the maximum permissible flying hours in a year (900hrs), this is equivalent to a rate of 1 report of involuntary sleep per 12,462,300 flying hours to the CAA vs 1 report of involuntary sleep in 137 flying hours via the 2-Way App adopted in this Study. Whilst a heavily approximated comparison, since among other factors, many pilots may not fly the total maximum permissible hours in a year, the considerable scale of this difference in reporting rates nonetheless highlights a fundamental flaw in relying on self-report to safety occurrence databases for incidence rates on in-flight fatigue-related incapacitation.

Beyond the accuracy of occurrence rates, concerns around the quality of information and data capture of pilot fatigue in formal aviation safety occurrence databases have also been discussed within prior research initiatives (Lyman & Orlady, 1981; Pouliquen et al., 2005). Supporting these prior findings, Study 1 identified poor reporting quality in content provided by the free text narratives in MORs, where there is no specified vernacular or mandatory fatigue risk data fields that reporters must fill out. Indeed, it seems that without a guiding framework for the collection of this safety information, insights on the sleepiness status of pilots or roster information is limited and quantitative analysis of specific fatigue risk data across time is not possible or may be unreliable. Study 2, in providing an optional free text narrative box within the 2-Way App, further evidenced this issue. The accompanying text to 1215 of the submitted ratings was not found to contain specific FRT information and was for most instances of supplied information too generic, colloquial or otherwise poor quality to be rigorously appraised. Hence for the purposes of increased fatigue risk management within aviation safety regulation, both at the level where safety incidents occur and as part of routine monitoring of alertness during normal operations, a set of common data requirements with respect to fatigue risk information appears critical alongside the practical and timely collation of such data.

Quality improvement opportunities for future reporting requirements

In safety research, there is undoubtedly considerable value in the collection of fatigue-related occurrence information which is expressed freely and idiosyncratically in the reporter's own descriptor terms. Safety forms which are exclusively proscriptive can easily devolve into tick box exercises where the organisational parsing up of information leads to disjointed data that is stripped of contextual meaning or nuance. As other safety researchers have warned, the danger of lobotomising data too heavily is that 'the big picture' behind the safety occurrence experienced by the reporter can be lost or even suppressed in the pursuit of efficient safety reporting processes and monitoring activities (Dekker, 2015). Too many constraints on safety reports may furthermore be off-putting or overly burdensome to reporters, and limit the potential of new risk markers to emerge or the overall quantity of data collected. Notwithstanding such caveats, the findings of both studies within this thesis suggest that most fatigue-related MOR narratives (Study 1) or free text commentary of information (Study 2) do not on their own provide details that are either sufficient or precise enough to enable a thorough analysis of fatigue factors or incident levels at either the individual report or aggregate dataset level. As such, for the purposes of safety reporting, it is strongly suggested that a common core of fatigue risk data metrics and accompanying roster information are adopted alongside any free text narratives within MORs. An industry-wide reporting framework structured with suggested metrics would enable future reporters to supply key data points relating to the distal, cumulative and proximal features of fatigue risk that may be relevant to their report, whilst permitting aggregate insights on the relationships between fatigue-related risks and aviation safety incursions. In addition, such a requirement would help identify for analysis MORs that may have elevated fatigue-risk distal indices, but have not been labelled or volunteered by the reporter as 'fatigue-related'. This would in turn allow future research studies to investigate fatigue risk trends in safety occurrence reports through additional objective methods, such as the bio-mathematical modelling of pilot rosters for all MORs submitted to the CAA, whether or not fatigue is explicitly cited in the MOR. Basic principles and suggested data requirements for such metrics are set out below;

1. Inclusion of precise timing data regarding duty schedules and sleep-wake history

The inclusion of precise timing and time zone data relating to the reporter's pre-duty work schedule and sleep-wake history is essential for assessing the likely contribution of homeostatic and circadian drivers to the reporter's fatigue state.

Rationale

The collection of precise timing information surrounding safety occurrences is critical for understanding the likelihood and severity of fatigue risk, and whether such risks are overlapping with critical phases of flight or other periods of elevated safety risk to the operation or individual. Such timing information is furthermore important to understand whether such risks represented *a priori* predictable schedule-driven or sleep-driven fatigue risks, or in fact were unexpected at the time the pilot started their duty, and arose through individual or operational 'on-the-day' factors. Timing data is furthermore important for establishing whether there are systemic trends in specific fatigue risk parameters associated with safety incursions. For example, time of day data has already been extensively shown in the road and rail industries to provide useful insights on the risk relationships between sleep and circadian parameters and daily prevalence peaks in road traffic and rail incidents and accidents (Eskandarian & Mortazavi, 2007; Horne & Reyner, 1995a; Pack et al., 1995). The mandatory requirement of work pattern data (including relevant circadian information such as time zone changes) would hence substantially improve the ability to track the specific working practices leading to severe levels of fatigue resulting in safety compromise from the pilot's perspective. Given widespread underreporting of fatigue within aviation, the collection of work pattern and individual specific sleep and circadian timing data in MORs is also important for the evaluation of objective fatigue related risks where self-disclosure has not been forthcoming. The collection of objective timing data potentially captures other safety reports which have not been labelled as 'fatigue related' by the reporter but nevertheless indicate elevated fatigue risks through objective analyses, such as biomathematical fatigue modelling of sleep and duty times.

2. Inclusion of a validated measure of sleepiness

The description of fatigue experience of the reporter is likely to be benefitted by the addition of a subjective rating of sleepiness on a scientifically validated scale denoting the severity of symptoms, and their period of duration.

Rationale

Self- ratings on validated sleepiness scales, whilst not holistic indications of someone's entire fatigue experience, can offer much greater clarity on the sleepiness component of the reporter's fatigue, the risk of sleep-driven performance decrement and more reliably interpreted by third party assessors. Hence such ratings offer greater comparative value on the severity of sleep-driven fatigue experienced by the reporter at the time of the event. The results from Study 1 suggested that symptoms relating to decreased alertness were cited in over half of the MORs in the study sample. Hence there would be greater organisational learning if more of these reports could be better understood and more systematically compared through the use of a validated sleepiness scale such as the KSS, denoting severity of symptoms.

3. Safety prompts for reports which do not have any proximal operational consequences

In order to assess the risks to human sub-system safety within the context of commercial aviation, the report should indicate whether both flying pilots were affected by fatigue and over what time period. In addition, the reporter should be encouraged to indicate whether their report relates to a work pattern, schedule feature that is likely to reoccur or any longer standing individual factors that may contribute to their in-flight fatigue.

Rationale

The reporter should be encouraged to explain why in their view, their fatigued state did not translate into an incident. Such information is important for understanding the role that other organisational or individual level safety defences may have in protecting overall system safety, on occasions where one or more flight crew are suffering from in-flight fatigue. In addition, reporters should highlight the specifics of existing schedule practices which may have led to the event, since such factors are unlikely to be unique and may represent underlying schedule-driven risks that have the potential to threaten future operations.

Appendix ii provides specific questions that may be considered as examples for these principles. Hence in order to better understand the relationships between sleep deficit, fatigue experience during flight and safety events in commercial aviation, more systematic collection of data during the MOR submission processes and retrieval of reporters' roster data within MORs is necessary. If basic roster details were routinely gathered within MORs and monitored from industry, this would permit research into the identification of any trends in reduced sleep opportunity or elevated KSS risk associated with serious occurrence data (whether explicitly reported as 'fatigue-related' or not). This would furthermore allow regulators and the broader aviation industry greater awareness of the predictable fatigue risks during normal operations that may be associated with serious occurrences – to compare against the baseline rates at which these roster-driven fatigue risks are expected to occur during normal operations.

5.3.2 Monitoring of fatigue experience (FRT 3 data)

Findings from Studies 1 and 2 suggested that post-flight fatigue reporting is by itself, methodologically insufficient for fatigue risk monitoring of in-flight sleepiness and sleep events within the aviation industry. There are potentially a wealth of factors that influence the reporting of fatigue or sleepiness related incapacitation via formal company or regulatory channels. However, from the stakeholder consultation process undertaken in Study 2, prominent barriers to fatigue reporting that were identified included the practical ease of reporting, pilots' perception that their fatigue report will not change anything, or fear of negative professional consequences following the submission of a fatigue report via a formal channel. Such reasons are consistent with survey findings conducted with hundreds of pilots from multiple European countries (European Cockpit Association, 2012; p3, p10), and reporting culture has been identified as an important research gap to be addressed within studies assessing fatigue risk management implementation (Bendak & Rashid, 2020). Against this context, effective data gathering on in-flight sleepiness events may need to rely more on the use of simple, quick app rating technology and provision of research monitoring exercises which in some way anonymize or confidentially control the reporter's identity in order to offer more accurate information on incidence rates. Indeed, as a quick and easy method of reporting, it is probable that the 2-Way app was imperative to enabling the monitoring of pilot sleepiness levels due to high levels of practical engagement. From a review of the published literature, it does not seem that a similar study has ever been conducted with participants from such a wide range of airlines over a month long period, and as such, these findings provide a new benchmark across industry. Since conducting Study 2, there has been one further large scale study collecting top of descent fatigue data with European pilots conducted over a two week period that was commissioned by the European Commission for assessment of long night and disruptive duties (Sallinen et al., 2020). This study also made use of a commercial application to facilitate collection of data, further highlighting the feasibility of this method of data collection on the flight deck, although submission rates and incidence levels were not provided. To the extent to which pro-active reporting of experience of in-flight fatigue is relied upon within the context of safety management systems and fatigue risk regulation, this thesis therefore highlights the ability and potential high desirability of increased use of practical app technology for pilots to self-report their levels of in-flight sleepiness at any stage of flight, where safe to do so, under conditions where some level of confidentiality to the reporter is guaranteed.

It is furthermore the case that app technology and guidance over the regular use of KSS for self-monitoring could be used across a number of different safety-critical industries, particularly where formal channels of fatigue reporting do not exist or staff may feel unable to highlight day-to-day fatigue concerns. Indeed, the ability of front-line remotely working staff to proactively rate their alertness levels through the 2-Way App as part of a research initiative was highly commended by the Trades Union Congress, since practical engagement and provision of data over fatigue from staff whilst at work is often challenging (TUC, 2018). Against this context, there is a real need to develop reliable proactive surveillance and analysis methodology that is both acceptable to front line staff and can more accurately provide data on fatigue exceedances or breaches to safe alertness alongside any reactive post-flight report systems that enable learning about the known contexts where operator fatigue may threaten safety. The methodology adopted in study 2 highlights the potential benefit of routine exercises using a quick user friendly app with a validated sleepiness measure in conjunction with bio-mathematical modelling of rosters in order to collect data during duty from a large ‘remote-working’ sample and compare fatigue experience of operators against expected fatigue risks.

However, there are still methodological limitations associated with this approach. As highlighted in Chapter 1, microsleeping events may occur outside of the pilots' awareness (e.g. Wright & McGown, 2001) and even if such events are identified, the findings of both studies suggest that there are clearly still additional cultural or practical barriers which act collectively to reduce the extent to which such events are reported via formal organizational channels. Hence, even during normal duties where safety events have not occurred, fatigue surveillance is limited where it is solely reliant on subjective reporting. As referred to in Chapter 2, passive monitoring technologies that provide real-time detection of observable proxies for sleepiness (such as measurement of blinks and eye-movements) methodologically have the potential to offer superior data capture on the pilots' fatigue status during flight (FRT 3), since such methods do not rely on the individual to both recall and report their symptoms of fatigue that may impact safe performance. There are clearly a range of scientific, practical, financial and industrial issues surrounding the deployment of such technologies during normal commercial flight operations, the debates for which are not the subject of direct enquiry for this thesis. However, the findings of both studies suggest that at a minimum, the scientific benefits should continue to be considered, particularly as there is growing precedent for such activities in other industries in the UK. For example, in 2016 a tram driver was found to have been micro-sleeping on duty which lead to the tram overspeeding and ultimately derailing, resulting in many injuries and the death of seven individuals (Rail Accident Investigation Branch (RAIB), 2020). The RAIB concluded that a serious systems level weakness was the lack of real-time mechanism to monitor driver alertness levels and provide mitigations for the impact to safe human performance. The use of modern technology to both detect and intervene where drivers 'lose awareness' of the driving task was hence a key recommendation in accident report, which the Croydon Tram link unit are currently trialing. Hence improvement in the monitoring of fatigue symptoms during duty (FRT 3 data) will likely require not only improvements in reporting practices but also consideration of new passive and non-intrusive modes of behavioural monitoring that may not only collect data but also serve as an alert to the affected operator.

5.3.3. The role of bio-mathematical modelling for assessment of fatigue risk regulatory principles

Biomathematical models of fatigue that incorporate the known temporal dynamics of sleep-wake regulation and circadian rhythms have been increasingly adopted as part of fatigue risk management practices in commercial aviation (Bendak & Rashid, 2020). Such model use has typically been deployed by companies to compare differences in relative risks for different duty schedules, provide guidance for the timing of use of fatigue countermeasures, aid post-hoc incident investigation of fatigue risks and help build safety cases for alternative means of regulatory compliance. However, the approach adopted in Study 2 of this thesis highlights the potential substantial value of normative research using bio-mathematical modelling of pilot rosters for regulators and safety researchers to monitor and track over time. This thesis provided research offering a baseline of predicted fatigue risk exposures across the UK aviation industry during a one month period. Such normative research, if conducted regularly throughout the year would enable regulators a greater evidence-base with which to make assessments over whether rostering practices meet or exceed limits on working hours, or are being interpreted in ways that reflect the intent of the performance-based regulations on fatigue. As discussed in Chapter 3, of the 10.5% of MORs which provided specific sleep lengths, the reported sleep achieved ranged between 0-6 hours sleep prior to duty. Whilst there was insufficient detail provided by the majority of MORs to provide clarity on the average predicted level of sleep deficit experienced by reporters from their rosters, the findings from Study 2 suggested such reduced sleep opportunities of less than six hours are not uncommon, and were predicted for a quarter of all flying duty periods in the field study conducted during normal operations. In this example, the findings from both studies suggest that the regulatory controls over schedule design emphasizing the requirement for at least an eight hour window of sleep for pilots between duty periods (European Aviation Safety Agency, 2014: ORO.FTL.235), appear to not be fully effective in their impact on roster construction. Hence, the use of such modelling on samples of rosters would allow for objective, independent assessments over the effectiveness of fatigue risk management practices with respect to different elements of aviation regulations. In addition, research initiatives on such data would enable hypotheses based on known sleep-driven fatigue risks (such as rosters containing reduced sleep opportunities) to be tested, and the relationships between schedule parameters and fatigue-related safety outcomes to be better understood and used to inform guidance.

As bio-mathematical models of sleepiness or ‘cognitive efficiency’ are continuing to be developed and validated in operational settings, it is important to discuss their current main limitations. In general, they appear less accurate for predicting the effects of chronic sleep restriction (Ramakrishnan et al., 2016). This seems to be for two main reasons. First, models do not always take into account sufficient prior sleep/wake history in formulating sleep and wake predictions (Mallis et al., 2004). This means that many programs do not calculate accumulating sleep loss over a sufficient period of days or even weeks in the model predictions of waking state. Second, they appear to inaccurately specify the relationship between the lower and upper asymptotes of the homeostatic sleep process (Johnson et al., 2004). Whilst recovery from total sleep deprivation may be predicted with reasonable accuracy, recovery following chronically restricted sleep is typically much slower in real life workers than would be predicted by these models (Ramakrishnan et al., 2016), although the success of validation efforts to address this may only be contained in commercially owned science (Hursh & van Dongen, 2011). Finally, most models are designed to produce average estimates of sleep timing and duration, and average sleepiness predictions (CASA, 2014). This is perhaps the most efficient design for the mass analysis of many schedules, and the assessment of average likely sleepiness levels during duty hours for a particular cohort. However, not all models provide confidence intervals associated with each predicted value (CASA, 2014), which means the likely range across individuals may not always be appreciated by model users. At the individual level, research into inter-individual differences in sleep and circadian processes clearly highlights the theoretical possibility of improving the accuracy of model predictions by taking into account specific individual traits and behaviours (van Dongen & Belenky, 2009). However, currently it is not clear how much additional predictive power such factors would bring to the baseline model predictions. For example, in Darwent and colleagues' study with train drivers using an early version of the SWP (Darwent et al., 2012), the addition of individual predictors such age, chronotype and experience to the model was not seen to make a statistically significant contribution to the sleep or sleepiness predictions. In addition, the addition of individual pilots' age and chronotype were not found to be significant additions to the SWP (two-process model) used in Study 2.

Perhaps one of the greatest limitations of bio-mathematical models of fatigue is that there are very few systematic, objective, detailed independent studies of the accuracy of predictions of any model relative to operational outcomes. As such, initiatives that attempt to investigate the models' ecological validity relative to different real-world scenarios should be a high priority, and model vendors should be encouraged to publish their programmes for research purposes. The SWP model is one of the only examples of where the full algorithms underpinning the modelling of circadian and homeostatic processes have been published (Åkerstedt & Folkard, 1996; Åkerstedt et al., 2008), as increasingly model development in the area of fatigue has become commercially sensitive and highly lucrative. The implication of this commercial context is that for some bio-mathematical modelling products, the key output parameters describing levels of risk and fatigue may have limited meaning to external scientists or be so loosely defined that they cannot be subject to scrutiny.

5.3.4. Fatigue risk data and the implications for FRMS

As discussed in sections 1.08- 1.10, Fatigue risk management systems are increasingly seen as important organizational controls regarding the hazard of fatigue. In theory FRMSs benefit from the collection of data via both prospective and retrospective methodologies, in what can be viewed as a hybrid of *Safety I* and *II* theoretical approaches. Under the FRMS framework, safety learning regarding fatigue is expected to not only depend on post-hoc safety reports of incidents or accidents, but rather to draw together data insights on both predicted fatigue risks and pilot experience of fatigue generated both prior to and during duty. Thus, far from being a bi-modal approach to safety, fatigue risk data can theoretically be collected across different points in time, different organisational layers (from rostering affecting the entire workforce, to individual pilot experience during duty) as well as across an increasing spectrum of severity and safety outcomes. In turn, such data is proposed to not exist in isolation, but rather to feed into an organisation's SMS, as one of the data streams relating to broader system safety. In practice however, in order to function effectively the data-driven element of FRMS is reliant on a) the adoption of consistent measures of fatigue risk, b) a robust reporting culture, c) agreement over threshold levels of fatigue risks and their relationships to foreseeable detriments in pilot functioning, and d) methods to relate the impact of such risks to broader system safety. The preceding sub-sections of 5.3 have set out some major implications for the measurement of fatigue risks and practical data collection techniques that should be considered for FRMS, building on the data categories derived from the FRT framework. However, for FRMS to generate organisational learning and ongoing safety improvement, an industry wide consensus needs to be achieved on fatigue threshold levels using consistent measures. Within an ultra-safe socio-technical system such as aviation, lack of consensus on fatigue thresholds impairs safety. It risks leading to an excessive focus on Safety I data around specific, though rare fatigue-related serious incidents and accidents that provide a highly misleading indicator of the safety implications of fatigue (see Figure 19, top right hand quadrant.)

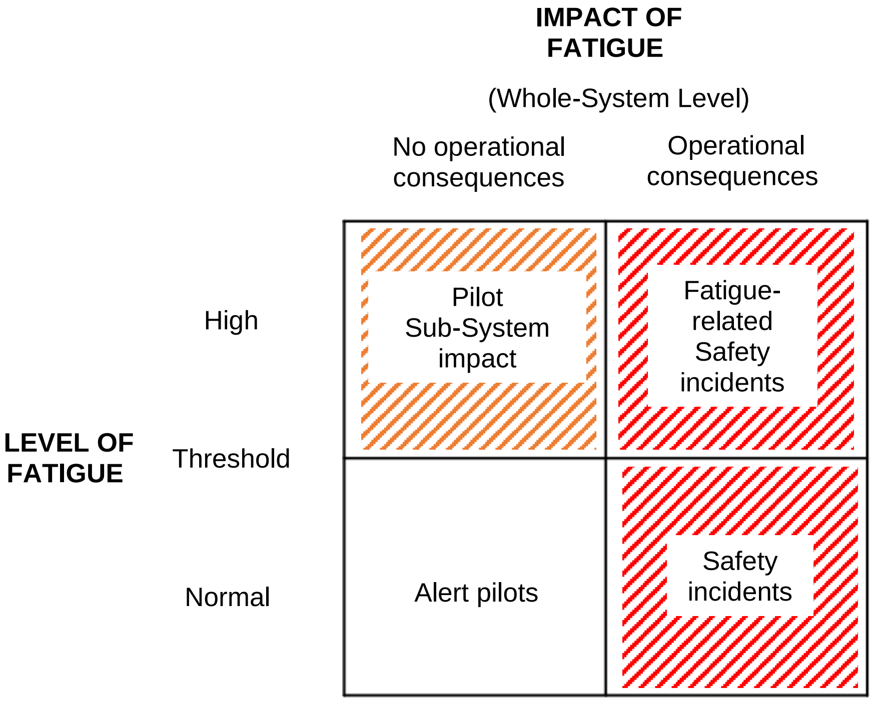


Figure 19. Data insights required by FRMS relating to the safety impact of fatigue

By contrast, with industry-wide agreement on fatigue threshold levels in place, independent assessments of the extent to which the invisible safety buffer of having ‘alert pilots’ in the cockpit is compromised and the frequency with which high fatigue levels are simultaneously affecting both pilots (endangering the pilot sub-system, even without operational consequences) can be systematically measured, appraised and comprehensively aggregated across industry. It is on such data insights (illustrated in Figure 19, top left hand quadrant) that this thesis has provided unique evidence and in Study 2 has established the first bench mark for the UK commercial aviation industry.

5.3.5. Achieving consistency of human factor safety standards in aviation

A further major conceptual development advanced in this thesis has been to specify a method to compare predicted fatigue risks and pilot experience (observed KSS ratings and involuntary sleep reports) with other system safety risks through applying a common flight hour denominator. In practice, such a move towards achieving consistent risk standards across the human factor component would require airline operators to accept the need to align their fatigue risk management practices with common standards set on an industry-wide basis. There would also be an ongoing need for independent research to inform on risk thresholds in light of emerging aggregated safety data from enhanced reporting and data inputs on the incidence and severity of in-flight sleepiness levels experienced by pilots.

Three prominent approaches for the development of safety standards within the transport sector may be summarized as 1) the evidential approach (adopting threshold levels informed by evidence-based points of risk inflection); 2) the systems reliability approach (where risk rates are developed to be consistent or in some important way aligned with other existing standards) and 3) the socio-political approach (where risks or the interventions to certain risks are considered against broader assessments and consultations regarding public acceptability). Such approaches are not mutually exclusive, but together provide useful context for ongoing appraisal of the practitioner question of 'how tired is too tired for safe flight?' and for evaluating the significance of normative fatigue risk data against the sleep scientific evidence and aviation safety standards literatures.

As discussed in greater detail in Chapters 2 and 4, the KSS is a scale of subjective sleepiness experience which has been the subject of substantial validation research with EEG variables (Åkerstedt et al., 2014; Kaida et al., 2006b; Kecklund & Åkerstedt, 1993; Sagaspe et al., 2008) and has been recognised within the applied field of sleep science as offering perhaps the best practical indications of evidence-based points of risk inflection based on sleep propensity. Thus use of the KSS metric to determine sleepiness severity provides an evidential basis for interpretation of scores for the upper end of the KSS scale (KSS 7-9) both in terms of neurobehavioural functioning and human performance decline (particularly where such performance is related to reaction times, vigilance and lapses in concentration). Against this context, the KSS findings from Study 2 suggest that UK commercial pilots experience high levels of sleepiness during their duties and are rostered duties that carry elevated fatigue risk exposure at levels which predictably exceed safe human performance.

However, it is also the case that the overall safety impact of any form of human performance failure (whether as a consequence of sleepiness or otherwise) is to a certain extent task and context dependent, and therefore likely to be substantially reduced in ultra-safe high reliability environments such as commercial aviation. Thus, even with evidence of severe sleepiness levels in pilots, it is clear that airplanes do not 'fall out of the sky' as an automatic consequence, as there is greater whole-system tolerance for human performance failures to occur without adverse operational consequences. Thus it is also helpful to consider a 'systems reliability' approach for appraising the tolerance of fatigue risk rates for the human sub-system. Where this comparison is made, it clear that sleepiness risks far exceed other forms of risk relating to the break down of human performance within the aviation setting. As outlined in Chapter 1, for medical incapacitations (which represent varying impact to pilot functioning via medical causes), there is a target rate of less than one occurrence per 1,000,000 hours. Study 2 found a schedule driven rate of 1 in 224 flight hours predicted at or above KSS 8 and a reported rate of 1 KSS at or above 8 per 59 hours and 1 report of involuntary sleep on the flight deck of one occurrence per 137 flight hours. These observed rates are likely an underestimate due to the further twenty reports discounted due to unavailability of full participant roster information, and the fact that pilots may be unaware of many involuntary sleep events. Notwithstanding this, compared to guidance recommending a target medical incapacitation rate of no more than 1 occurrence per 1000,000 hours, the predicted and observed fatigue risk rates arguably represent non-trivial differences in risk rates. It should be noted that although difficult to measure, there is evidence that the medical incapacitation target rate is also exceeded by real world medical incapacitations, with a national estimate of 2.5 occurrences per 1000,000 hours (Evans & Radcliffe, 2012). However, from a systems reliability perspective, this finding does not discount the fact that there appears to be a fundamental disparity between the acceptable probabilities of observed and schedule-driven on-duty fatigue risks and medical incapacitation risks, despite both risk rates relating to significant in-flight functional impairment of crew. As such, this research provides new data to inform safety-related decision making. Since the target medical incapacitation rate was based on the rationale of a single pilot incapacitation in a multi-crew environment, a further important finding was that on four occasions during Study 2, involuntary sleep was reported for both flight crew during the same flight, equating to a rate of possible overlapping involuntary sleep events and simultaneous sleepiness in both flight crew of 1 per 1818 flight hours, which again is far higher than the observed or target medical incapacitation rates. Simultaneous sleepiness within both flight crew is of elevated safety concern since it represents a form of common mode failure of a safety critical system (Downer, 2009). The assurance otherwise provided by having independent risk of failure from multiple crew may be compromised where sleepiness risks are similar for pilots flying the same work schedules. As such, the observed rate of simultaneous sleep-related events on the flight deck appears very high against the medical incapacitation rate which is based on single pilot incapacitation. Thus such findings offer a useful basis for comparison between current acceptable risk standards with respect to the break down in pilot functioning and performance impairment during flight from different causes, and should prompt further discussions between sleep-science and medical practitioners involved in aviation safety. This is particularly since the aeromedical regulatory process evolves over time in light of new evidence of in-flight impairments that are affecting pilots (ATSB, 2007). As such, it seems important that continued research into pilot fatigue and sleepiness events during actual flying hours are monitored against similar risk standards to other forms of in-flight impairment and incapacitations.

A further standard relating to the acceptability of risk to safe pilot performance is that concerning alcohol-related impairment. Under both EU regulations (EU 965/2012 CAT.GEN.MPA.100 (c)) and UK regulations subsequently retained and applied from 14th August 2018 (UK Reg EU 956/2012; EU 2018/1042), flight crew members are not permitted to perform duties on an aircraft when under the influence of alcohol, and the threshold level of tolerance is a blood alcohol concentration (BAC) of 0.02%. Previous research efforts to provide industries and regulators with an understandable index of relative risks associated with fatigue have sought to equate the performance impairment of sleep loss and alcohol intoxication (Dawson & Reid, 1997; Lamond & Dawson, 2002; Maruff et al., 2005; Williamson & Feyer, 2000). As discussed in Chapter 1, there are some differences between the findings and interpretation of these studies (particularly with respect to the equivalence of magnitude of performance decline). However, broadly speaking these studies suggest that cognitive impairments increase within individuals who are awake continuously beyond sixteen hours in quantitatively and qualitatively similar ways to increases in BAC %. In Maruff's study which applied more detailed statistical measures, the equivalence of this impairment could mainly be seen through performance variability (reaching equivalence to 0.02% BAC by 14 hours, 0.05% BAC by 24 hours). Given suggestive evidence of these markers of performance equivalence, it is perhaps noteworthy that pilots would be prosecuted and face up to two years in prison if found in breach of the 0.02 BAC % threshold (Railways and Transport Act, 2003). However, evidence from this PhD and from other prior studies has shown that commercial pilots can exceed sixteen hours of continual wakefulness whilst at work or may even face duty scenarios that result in their being continually awake beyond 19 hours (Gander et al., 1998). Whilst there are clearly limitations to the comparisons between the effect of alcohol and sleep loss and the methods used to compare the two; there is nevertheless strong evidence that both induce increasing performance variability, and some evidence that even mild forms of sleep loss (<17 hours) may result in equivalence to or exceed the acceptable BAC % for commercial aviation.

Together these findings highlight the need for consistent risk standards to be applied across the human component to avoid a weak link in safety defences; they also raise important questions on whether pilot fatigue should continue to be regulated differently to other human factors affecting safe pilot performance. Indeed, from a safety perspective, sleep loss and fatigue can substantially compromise the pilot's functional state and behaviour, and the risk of simultaneous in-flight fatigue is not necessarily independent between flight crew undertaking the same duties. Thus it is important that continued evidence-based research on pilot fatigue enables comparison with other human factors risks where greater clarity on threshold levels of acceptability have already been established. This PhD has also offered a baseline of the severity and exposure levels of fatigue and sleepiness events across the majority of UK airlines using the same metrics, and has provided per-flight hour rates that facilitate industry-specific discussions regarding risk. These data hence provide new opportunities for discussions with relevant industry stakeholders with respect to the question of 'how tired is too tired for safe flight?'.

5.4 Further research

In the following section, opportunities for research to help contribute further evidence fall under two areas;

1) Research that addresses the limitations of Studies 1 and 2
2) Research that provides further evidence regarding causal relationships between sleep loss, pilot performance and safe flight

The findings from Study 1 highlight a need for improvement for the quality of reporting fatigue risk information and this chapter sets out a suggested framework for the CAA to consider mandating to achieve substantial data quality improvements. A repeat analysis of the fatigue-related MOR dataset once regulatory efforts have been deployed to improve reporting metrics would hence be an important future aspiration. However, further research that could be undertaken immediately would be to conduct an anonymous survey with reporters to the MOR system to help to elucidate any additional barriers to data collection from pilots. Certain barriers to general fatigue reporting from pilots to their own companies were highlighted during the stakeholder consultation in Study 2 which included the finding that many pilots find fatigue reporting forms laborious to fill out, and fear negative professional consequences from submitting fatigue reports. It would therefore be helpful for future projects looking to improve or redesign the MOR system to address precisely which barriers are perceived to have the greatest impact on reporting rates and the quality of data submitted and whether these could be addressed through technological or organisational solutions. For example, questions could attempt to elucidate the extent to which pilots feel jeopardy in their submissions that may bias the inclusion or exclusion of certain risk information. It would also be important to understand why some reports to the CAA are transposed into third hand by the pilot company's organisation whilst others are not, since important safety-related information may be lost and misinterpreted during this process.

There are a number of different ways in which the findings from Study 2 could be built upon. The first would be to conduct research with volunteer airlines that could complement the fatigue prediction (FRT 1 – roster analysis data) and sleepiness measurement (FRT 3 - KSS data collection) protocol adopted in Study 2 with the analysis of concurrent flight data monitoring data (FRT 4 - operational performance data). Flight data monitoring systems transmit data pertaining to a range of aircraft parameters retrieved from internal and external sensors and recorders of the aircraft (such as the flight data recorder and cockpit voice recorder). This means that aspects of the aircraft trajectory (e.g. speed, acceleration, altitude, pitch) and mode selections from the pilot (e.g. flap position, selected altitude) are captured every flight. Future research on the impact of sleepiness on flight performance would hence benefit from large scale long term exploratory analysis of high sleepiness levels in flight and trends in event monitoring derived from flight data. Particular performance deficits or events highlighted in the MOR data in Study 1 such as delayed or missed communications with Air Traffic Control or incorrect flap selections could also be explored to assess whether they represent potential system markers of pilot fatigue-related decline. A limitation of Study 2 was that participants were not mandated to submit ratings at particular operational points during flight (such as top of climb, top of descent etc). This was in part due the fact that airlines may have different safety protocols around operational points and as such, researchers would need to work closely with individual safety departments to establish a consistent set of rating points. However, for such future research incorporating flight data monitoring data, there would be substantial benefit in encouraging the consistent submission of ratings at key operational points in order to better understand the relationships between alertness/ sleepiness profiles and pilot interactions with the flight deck and Air Traffic Control.

The focus of Study 2 was to understand fatigue risk exposure and sleepiness rates during flight. However, a key area for future research would be to explore in greater depth the profiles of alertness in pilots following their flight duties and the associated time course of recovery for pilots to return to feeling alert during their out of work hours or days off in-between roster blocks. Two prominent themes identified in Study 1 MOR reports related to the reporter's insufficient pre-duty sleep and cumulative fatigue built up from consecutive sleep loss over several duties. Whilst some laboratory studies have already highlighted that recovery from acute total sleep deprivation differs from chronic partial sleep deprivation, the typical study period does not exceed two weeks. Hence, the protocol from Study 2, if conducted over longer periods of time and encouraging multiple KSS ratings during days off could reveal greater real world insights on the time course of recovery profiles from different schedules and forms of sleep debt. These data are particularly important to consider for fatigue prediction efforts, since a key limitation associated with a number of bio-mathematical fatigue models is that they do not appear to adjust predictions to account for cumulative sleep loss over long time periods. Indeed, this limitation is supported by the finding that KSS predictions from the two-process model were significantly improved by the addition of variables that indexed consecutive early or late duties in the prior 72 hours in Study 2.

There has also been an increasing research interest in the impact of duty schedules and pilot fatigue with respect to broader health and safety consequences. Some recent survey research on the links between stress, physical health and pilot fatigue suggests that just under 20% of Australian and European pilots have reported significant depression symptoms (Venus & Holforth, 2022) and greater psychosocial stress is significantly correlated with sleep problems and higher fatigue levels in pilots (Venus & Holtforth, 2021). Whilst intercorrelations between stress, sleep difficulties, fatigue and health have previously been reported in a number of studies (Bennett, 2003; lo Martire et al., 2020; Sneddon et al., 2013; van Reeth et al., 2000) more research is needed to assess how such factors may be exacerbated or reduced by pilot work schedules. Thus further research adopting a similar method to Study 2 that assesses in greater detail how duty parameters may influence not only in-flight sleepiness levels but also stress and health outcomes over both short and longer term periods of time would be highly beneficial, and build on these latest high level survey insights.

A further way in which the findings from Study 2 could be built upon would be to conduct a repetition of the protocol during different times of year. It should be noted that it is likely that there is some ‘seasonality’ impact on rosters (e.g. denoting greater volume of flights, more demanding schedules) during the summer season which represents greater fatigue risk exposure than other times of year, particularly for short haul pilots, and hence this would be important to explore through follow up replication studies to provide relative context to the exposure findings of Study 2. Moreover, since Study 2 was conducted, there has been a substantial impact to the aviation industry through reduction in operations and staffing during the COVID-19 pandemic. As such, as the industry continues to recover and return to normal operations, a replication of Study 2 appears important for assessing the current fatigue risk exposure associated with UK commercial aviation operations.

For the investigation of causal relationships between of sleep loss, sleepiness and safe pilot performance, an experimental research approach was first considered and developed at the outset of this PhD. The approach proposed an experimental study using a commercial aircraft simulator which would seek to investigate the association between increasing levels of controlled sleep deprivation (FRT 2), pilot neurobehavioural signs and symptoms (FRT 3) and flight performance (FRT 4) during safety-critical scenarios. The methodological benefit of an experimental simulator study is that it provides the most direct evidence on the impact of increasing sleepiness levels on pilot performance, that can be conducted in a safe, high - fidelity environment, with advanced performance monitoring opportunities. Within the field of applied sleep science, the precedent for combining a systematic sleep deprivation protocol with performance assessment in a simulated environment has been robustly demonstrated within the driving setting, and has led to substantial insights on the characteristics of fatigue-related performance decline in drivers and the relationship between increasing sleepiness and accident risk (Liu et al., 2009). Only a handful of studies have sought to investigate the impact of sleep-deprivation on piloting performance within the simulated aviation setting, with most either investigating extreme levels of sleep deprivation, or involving single pilot missions with military personnel (e.g. fighter pilot performance) which differ considerably to the nature of flying performance parameters typical in multi-crew commercial operations (Caldwell et al., 2004; Caldwell, Jr. et al., 2004; Leino et al., 2007; Neri et al., 1992; Russo et al., 2004). In terms of published research, there are no known simulator studies that have sought to systematically investigate the impact of more moderate forms of sleep deprivation on commercial pilot performance. Hence an important gap in the literature exists for a study of this type. Unfortunately, despite these clear methodological benefits, this approach was not able to be advanced for this PhD due to financial and practical constraints over simulator and pilot participant access using a systematic sleep deprivation protocol. Alternative methodological variations of this approach were considered to try to improve the feasibility of the study, such as attaching the experimental protocol onto existing simulator training sessions or pilots' proficiency examinations, from which an opportunistic sample of volunteer pilots could be drawn. Both the LPC (annual recurrent training) and OPC (semi-annual recurrent training) for pilots consist of training checks conducted in the full-flight simulator. However, these approaches contained substantial methodological limitations (e.g. reliance on natural variation in fatigue levels across pilots during their training; reduced simulator time available for the experimental protocol post training or exam) or prominent ethical considerations (e.g. if the pilot has a bad experience in a study as part of their simulator training experience there could be negative psychological impact that extends beyond the study or bears consequences to the pilots' licence). Future work in this area may hence be difficult to achieve, but methodologically would provide important new insights on the impact of pilot sleepiness on safe multi-crew performance akin to the simulator work initially undertaken to inform the medical incapacitation rate.

Conclusion

Operator fatigue is a leading safety concern in many shift-work settings. However, as an ultra-safe high risk industry, the accident risk profile of commercial aviation remains one of extremely low frequency of occurrence but exceptionally high severity in terms of human and financial costs. As such, the legal requirements for pilots to operate 'sufficiently free from fatigue' such as not to endanger safe flight raise important prediction, measurement and risk appraisal questions for industry stakeholders. In particular, such requirements emphasize the need for consensus on meaningful metrics and threshold levels of acceptability of fatigue risk exposure that are informed by scientific evidence and real world data. Using the Fatigue Risk Trajectory as an overarching conceptual framework, this thesis has structured a practical approach to provide new research insights on fatigue risk data from both adverse safety occurrences and normal commercial aviation operations. Study 1 highlighted the current limitations of the MOR safety occurrence database in capturing sufficient fatigue risk data to enable quantitative analysis of relationships between fatigue risk parameters and safety incursions. Qualitative assessment of fatigue risk themes raised in the reports however suggested considerable overlap in duty and sleep themes underscoring the potential value of the future systematic collection of data. Study 2 was able to address some of these data limitations in the live commercial aviation environment by capturing predicted and actual fatigue risk exposure levels during normal operations using consistent metrics, the severity levels and rates of which could be assessed against other safety threshold levels governing human performance in aviation. This research provided a first benchmark of sleepiness rates of occurrence with pilots from eighteen UK airlines across a month long period. Biomathematical model analyses indicated that a substantial proportion of flown pilot working schedules are likely to be associated with insufficient sleep opportunities prior to flying duties. High levels of in-flight sleepiness were both reported by pilots and predicted by biomathematical model estimations using the pilots' work schedule times as inputs. Such findings hence have important broader implications for assessing the efficacy of regulatory controls in limiting fatigue hazards in commercial aviation. Moreover, such findings were also appraised against sleep science and relevant existing safety thresholds for the human component in commercial aviation. Where such comparisons are made, there appear to be considerable differences in risk tolerance for high levels of fatigue, compared with other forms of potential pilot incapacitation, that are worthy of further consideration. Taken together, the findings from this thesis address important aspects of fatigue risk monitoring and have demonstrated the scope for generating new insights from both normal operations and when safety events occur.

References

Aaronson, L. S., Teel, C.S., Cassmeyer, V., Neuberger, G. B., Pallikkathayil, L., Pierce, J., Press, A. N., Williams, P. D., & Wingate, A. (1999). Defining and Measuring Fatigue. *Journal of Nursing Scholarship*, *31*(1), 45–50.

Abe, T., Nonomura, T., Komada, Y., Asoka, S., Sasai, T., Ueno, A., & Inoue, Y. (2011). Detecting deteriorated vigilance using percentage of eyelid closure time during behavioural maintenance of wakfulness tests. *International Journal of Psychophysiology*, *82*, 269–274.

Achermann, P., & Borbély, A. A. (2011). Sleep homeostasis and models of sleep regulation. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and Practice of Sleep Medicine* (5th ed., pp. 431–444). Elsevier Saunders.

Adan, A., & Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: A reduced scale. *Personality and Individual Differences, 12*(3), 241–253. https://doi.org/10.1016/0191-8869(91)90110-W

Åkerstedt, T. (1995). Work hours, sleepiness and the underlying mechanisms. *Journal of Sleep Research*, *4*(Suppl. 2), 15–22.

Åkerstedt, T. (2003). Shift work and disturbed sleep/wakefulness. In *Occupational Medicine* (Vol. 53, Issue 2, pp. 89–94). https://doi.org/10.1093/occmed/kqg046

Åkerstedt, T., & Gillberg, M. (1982). Displacement of the sleep period and sleep deprivation. *Human Neurobiology*, *1*, 163–171.

Åkerstedt, T., Knutsson, A., Westerholm, P., Theorell, T., Alfredsson, L., & Kecklund, G. (2002). Sleep disturbances, work stress and work hours A cross-sectional study. *Journal of Psychosomatic Research*, *53*, 741–748.

Åkerstedt, T., Anund, A., Axelsson, J., & Kecklund, G. (2014). Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *Journal of Sleep Research*, *23*(3), 240–252. https://doi.org/10.1111/jsr.12158

Åkerstedt, T., Björn, P., Anund, A., & Kecklund, G. (2005). Impaired alertness and performance driving home from the night shift: a driving simulator study. *Journal of Sleep Research*, *14*(1), 17–20. https://doi.org/10.1111/j.1365-2869.2004.00437.x

Åkerstedt, T., Connor, J., & Gray, A. (2008). Predicting road crashes from a mathematical model of alertness regulation — The Sleep / Wake Predictor. *Accident Analysis and Prevention*, *40*, 1480–1485. https://doi.org/10.1016/j.aap.2008.03.016

Åkerstedt, T., & Folkard, S. (1996). Predicting duration of sleep from the three process model of regulation of alertness. *Occupational and Environmental Medicine*, *53*(2), 136–141. https://doi.org/10.1136/oem.53.2.136

Åkerstedt, T., & Gillberg, M. (1990). Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*, *52*(1–2), 29–37. https://doi.org/10.3109/00207459008994241

Åkerstedt, T., Hallvig, D., Anund, A., Fors, C., Schwarz, J., & Kecklund, G. (2013). Having to stop driving at night because of dangerous sleepiness - awareness, physiology and behaviour. *Journal of Sleep Research*, *22*, 380-388. https://doi.org/10.1111/jsr.12042

Åkerstedt, T., Hallvig, D., & Kecklund, G. (2017). Normative data on the diurnal pattern of the Karolinska Sleepiness Scale ratings and its relation to age, sex, work, stress, sleep quality and sickness absence/illness in a large sample of daytime workers. *Journal of Sleep Research*, *26*(5), 559–566. https://doi.org/10.1111/jsr.12528

Åkerstedt, T., Ingre, M., Broman, J., & Kecklund, G. (2008). Disturbed Sleep in Shift Workers, Day Workers, and Insomniacs. *Chronobiology International*, *25*(2–3), 333–348. https://doi.org/10.1080/07420520802113922

Åkerstedt, T., Ingre, M., Kecklund, G., Anund, A., Sandberg, D., Wahde, M., Philip, P., & Kronberg, P. (2010). Reaction of sleepiness indicators to partial sleep deprivation, time of day and time on task in a driving simulator - the DROWSI project. *Journal of Sleep Research*, *19*(2), 298–309. https://doi.org/10.1111/j.1365-2869.2009.00796.x

Åkerstedt, T., Kecklund, G., & Axelsson, J. (2008). Effects of context on sleepiness self-ratings during repeated partial sleep deprivation. *Chronobiology International*, *25*(2-3), 271-278. https://doi.org/10.1080/07420520802110589

Åkerstedt, T., Kecklund, G., & Selén, J. (2010). Early Morning Work - Prevalence and relation to sleep/wake problems: A National Representative survey. *Chronobiology International*, *27*(5), 975–986. https://doi.org/10.3109/07420528.2010.489001

Åkerstedt, T., Knutsson, A., Westerholm, P., Theorell, T., Alfredsson, L., & Kecklund, G. (2004). Mental fatigue, work and sleep. *Journal of Psychosomatic Research*, *57*(5), 427–433. https://doi.org/10.1016/j.jpsychores.2003.12.001

Åkerstedt, T., Nilsson, P. M., & Kecklund, G. (2009). Sleep and recovery. In P. L. Perrewé & D. C. Ganster (Eds.), *Current Perspectives on Job-Stress Recovery* (Vol. 7, pp. 205–247). Emerald Group Publishing Limited. https://doi.org/10.1108/S1479-3555(2009)0000007009

Åkerstedt, T., & Torsvall, L. (1981). Shift work Shift-dependent well-being and individual differences. *Ergonomics*, *24*(4), 265–273. https://doi.org/10.1080/00140138108924850

Åkerstedt, T., & Wright, K. P. (2009). Sleep Loss and fatigue in Shift Work and Shift Work Disorder. *Sleep Medicine Clinics*, *4*(2), 257–271. https://doi.org/10.1016/j.jsmc.2009.03.001.Sleep

Alhola, P., & Polo-Kantola, P. (2007). Sleep deprivation: Impact on cognitive performance. *Neuropsychiatric Disease and Treatment*, *3*(5), 553–567.

Aljurf, T. M., Olaish, A. H., & BaHammam, A. S. (2018). Assessment of sleepiness, fatigue, and depression among Gulf Cooperation Council commercial airline pilots. *Sleep and Breathing*, *22*(2), 411–419. https://doi.org/10.1007/s11325-017-1565-7

Alqurashi, Y., Dawidziuk, A., Alqarni, A., Kelly, J., Moss, J., Polkey, M., & Morrell, M. (2021). A visual analog scale for the assessment of mild sleepiness in patients with obstructive sleep apnea and healthy participants. *Annals of Thoracic Medicine*, *16*(2), 141–147. https://doi.org/10.4103/atm.ATM_437_20

Alvaro, P. K., Jackson, M. L., Berlowitz, D. J., Swann, P., & Howard, M. E. (2016). Prolonged eyelid closure episodes during sleep deprivation in professional drivers. *Journal of Clinical Sleep Medicine*, *12*(8), 1099–1103. https://doi.org/10.5664/jcsm.6044

Amalberti, R. (2001). The paradoxes of almost totally safe transportation systems. *Safety Science*, *37*(2–3), 109–126. https://doi.org/10.1016/S0925-7535(00)00045-X

Amundsen, A. H., & Sagberg, F. (2003). *Hours of service regulations and the risk of fatigue- and sleep-related road accidents: a literature review*. Transportøkonomisk institutt (Norway & Sweden). Vägverket: Institute of Transport Economics.

Anund, A., Kecklund, G., Kircher, A., Tapani, A., & Åkerstedt, T. (2009). The Effects of Driving Situation on Sleepiness Indicators after Sleep Loss: A Driving Simulator Study. *Industrial Health*, *47*, 393–401.

Arendt, J. (2010). Shift work: Coping with the biological clock. *Occupational Medicine*, *60*(1), 10–20. https://doi.org/10.1093/occmed/kqp162

Arendt, J., & Middleton, B. (2018). Human seasonal and circadian studies in Antarctica. *General and Comparative Endocrinology*, *258*, 250–258.

Arendt, J., Middleton, B., Williams, P., Francis, G., & Luke, C. (2006). Sleep and circadian phase in a ship’s crew. *Journal of Biological Rhythms*, *21*(3), 214–221. https://doi.org/10.1177/0748730405285278

Arnedt, J.T., Wilde, G.J.S., Munt, P.W., & Maclean, A.W. (2000). Simulated driving performance following prolonged wakefulness and alcohol consumption: separate and combined contributions to impairment. *Journal of Sleep Research*, *9*, 233–241.

Australian Transport Safety Bureau. (2007). Pilot Incapacitation: Analysis of Medical Conditions Affecting Pilots Involved in Accidents and Incidents. *ATSB Transport Safety Report*, *B2006/0170*.

Aviation Confidential Incident Reporting Programme (CHIRP). (2017). *Air Transport Feedback Issue 121*. Retrieved: https://www.chirp.co.uk/upload/docs/Air Transport/ATFB 121 - January 2017.pdf

Axelsson, J., Åkerstedt, T., Kecklund, G., & Lowden, A. (2004). Tolerance to shift work - How does it relate to sleep and wakefulness? *International Archives of Occupational and Environmental Health*, *77*(2), 121–129. https://doi.org/10.1007/s00420-003-0482-1

Babcoff, H., Zukerman, G., Fostick, L., & Benartzi, E. (2005). Effect of the diurnal rhythm and 24 h of sleep deprivation on dichotic temporal order judgment. *Journal of Sleep Research*, *14*(1), 7–15. https://doi.org/10.1111/j.1365-2869.2004.00423.x

Baglioni, C., Spiegelhalder, K., Lombardo, C., & Riemann, D. (2010). Sleep and emotions: A focus on insomnia. *Sleep Medicine Reviews*, *14*(4), 227–238. https://doi.org/10.1016/j.smrv.2009.10.007

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

Balandong, R.P., Ahmad, R.F., Mohamad Saad, M.N., & Malik, A.S. (2018). A Review on EEG-Based Automatic Sleepiness Detection Systems for Driver. *IEEE Access: Special Section on Wearable and Implantable Devices and Systems*, *6*, 22908–22919. https://doi.org/10.1109/ACCESS.2018.2811723

Balkin, T. J., Rupp, T., Picchioni, D., & Wesensten, N. J. (2008). Sleep Loss and Sleepiness. *Chest*, *134*(3), 653–660. https://doi.org/10.1378/chest.08-1064

Banks, S., & Dinges, D. F. (2007). Behavioral and Physiological Consequences of Sleep Restriction. *Journal of Clinical Sleep Medicine*, *3*(5), 519-528

Barclay, N. L., Eley, T. C., Parsons, M. J., Willis, T. A., & Gregory, A. M. (2013). Monozygotic twin differences in non-shared environmental factors associated with chronotype. *Journal of Biological Rhythms*, *28*(1), 51–61.

Barger, L. K., Lockley, S. W., Rajaratnam, S. M. W., & Landrigan, C. P. (2009). Neurobehavioral, Health, and Safety Consequences Associated With Shift Work in

Safety-Sensitive Professions. *Current Neurology and Neuroscience Reports*, *9*(2), 155-164. doi: 10.1007/s11910-009-0024-7.

Barsevick, A., Frost, M., Zwinderman, A., Hall, P., & Halyard, M. (2010). I'm so tired: Biological and genetic mechanisms of cancer-related fatigue. *Quality of Life Research*, *19*(10), 1419–1427. https://doi.org/10.1007/s11136-010-9757-7

Bartlett, F. C. (1953). Psychological Criteria for Fatigue. In W. F. Floyd & A. T. Welford (Eds.), *Symposium on Fatigue* (pp. 1–5). H.K. Lewis.

Bartley, S. H., & Chute, E. (1947). *Fatigue and impairment in man.* McGraw-Hill Book Company. https://doi.org/10.1037/11772-000

Basheer, R., Strecker, R. E., Thakkar, M. M., & McCarley, R. W. (2004). Adenosine and sleep–wake regulation. *Progress in Neurobiology*, *73*(6), 379–396. https://doi.org/10.1016/j.pneurobio.2004.06.004

Basner, M., Mollicone, D., & Dinges, D. F. (2011). Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronautica*, *69*(11–12), 949–959. https://doi.org/10.1016/j.actaastro.2011.07.015

Basner, M., Rao, H., Goel, N., & Dinges, D. F. (2013). Sleep deprivation and neurobehavioral dynamics. *Current Opinion in Neurobiology*, *23*(5), 854–863. https://doi.org/10.1016/j.conb.2013.02.008

Beebe, D. W., & Gozal, D. (2002). Obstructive sleep apnea and the prefrontal cortex: towards a comprehensive model linking nocturnal upper airway obstruction to daytime cognitive and behavioral deficits. *Journal of Sleep Research*, *11*(1), 1–16.

Belenky, G., Wesensten, N. J., Thorne, D. R., Thomas, M. L., Sing, H. C., Redmond, D. P., Russo, M., & Balkin, T. J. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *Journal of Sleep Research*, *12*, 1–12.

Bendak, S., & Rashid, H. S. J. (2020). Fatigue in aviation: A systematic review of the literature. In *International Journal of Industrial Ergonomics* (Vol. 76). Elsevier B.V. https://doi.org/10.1016/j.ergon.2020.102928

Benington, J. H., & Craig Heller, H. (1995). Restoration of brain energy metabolism as the function of sleep. *Progress in Neurobiology*, *45*(4), 347–360. https://doi.org/10.1016/0301-0082(94)00057-O

Bennett, S. A. (2003). Flight crew stress and fatigue in low-cost commercial air operations - an appraisal. *International Journal of Risk Assessment and Management*, *4*(3), 207–231.

Binks, P. G., Waters, W. F., & Hurry, M. (1999). Short-Term Total Sleep Deprivations Does Not Selectively Impair Higher Cortical Functioning. *Sleep*, *22*(3), 328–334. https://doi.org/10.1093/sleep/22.3.328

Bjorvatn, B., Stangenes, K., Øyane, N., Forberg, K., Lowden, A., Holsten, F., & Åkerstedt, T. (2006). Subjective and Objective Measures of Adaptation and Readaptation to Night Work on an Oil Rig in the North Sea. *Sleep*, *29*(6), 821–829. https://doi.org/10.1093/sleep/29.6.821

Bonnefond, A., Härmä, M., Hakola, T., Sallinen, M., Kandolin, I., & Virkkala, J. (2006). Interaction of age with shift-related sleep-wakefulness, sleepiness, performance, and social life. *Experimental Aging Research*, *32*(2), 185–208. https://doi.org/10.1080/03610730600553968

Borbély, A. A. (1982). A Two Process Model of Sleep Regulation. *Human Neurobiology*, *1*, 195–204. https://doi.org/10.1111/jsr.12371

Borbély, A. A., & Achermann, P. (1992). Concepts and models of sleep regulation: An overview. *Journal of Sleep Research*, *1*(2), 63–79.

Borbély, A. A., Baumann, F., Brandeis, D., Strauch, I., & Lehmann, D. (1981). Sleep deprivation: Effect on sleep stages and EEG power density in man. *Electroencephalography and Clinical Neurophysiology*, *51*(5), 483–493. https://doi.org/10.1016/0013-4694(81)90225-X

Bostock, S., & Steptoe, A. (2013). Influences of early shift work on the diurnal cortisol rhythm, mood and sleep: Within-subject variation in male airline pilots. *Psychoneuroendocrinology*, *38*(4), 533–541. https://doi.org/10.1016/j.psyneuen.2012.07.012

Boudreau, P., Dumont, G. A., & Boivin, D. B. (2013). Circadian Adaptation to Night Shift Work Influences Sleep, Performance, Mood and the Autonomic Modulation of the Heart. *PLoS ONE*, *8*(7), e70813. https://doi.org/10.1371/journal.pone.0070813

Bourgeois-Bougrine, S., Carbon, P., Gounelle, C., Mollard, R., & Coblentz, A. (2003). Perceived fatigue for short- and long-haul flights: A survey of 739 airline pilots. *Aviation Space and Environmental Medicine*, *74*(10), 1072–1077.

Brendel, D. H., Reynolds, C. F., Jennings, J. R., Hoch, C. C., Monk, T. H., Berman, S. R., Hall, F. T., Buysse, D. J., & Kupfer, D. J. (1990). Sleep Stage Physiology, Mood, and Vigilance Responses to Total Sleep Deprivation in Healthy 80-Year-Olds and 20-Year-Olds. *Psychophysiology*, *27*(6), 677–685. https://doi.org/10.1111/j.1469-8986.1990.tb03193.x

British Broadcasting Corporation (BBC). (2013). *Jet pilots sleep on duty*. Retrieved 07/08/18: https://www.bbc.co.uk/news/av/uk-england-london-22059560/jet-pilots-sleep-on-duty-twice.

Brown, I. D. (1994). Driver Fatigue. *Human Factors*, *36*(2), 298–314. https://journals.sagepub.com/doi/10.1177/001872089403600210

Cabon, P., Bourgeois-Bougrine, S., Mollard, R., Coblentz, A., & Speyer, J. J. (2003). Electronic Pilot-Activity Monitor: A countermeasure against fatigue on long-haul flights. *Aviation Space and Environmental Medicine*, *74*, 679–682.

Cabon, P. H., Coblentz, A., Mollard, R., & Fouillot, J. P. (1993). Human vigilance in railway and long-haul flight operation. *Ergonomics*, *36*(9), 1019–1033. https://doi.org/10.1080/00140139308967974

Cabon, P., Mollard, R., Debouck, F., Chaudron, L., Gran, J.Y. & Deharvengt, S. (2008). From Flight Time Limitations to Fatigue Risk Management Systems. In E. Rigaud, E. Hollnagel & F. Pieri (Eds.), *Proceedings of the Third Resilience Engineering Symposium* (pp 27-33). Presses des Mines

Cajochen, C., Brunner, D., Krauchi, K., Graw, P., & Wirz-Justice, A. (1995). Power Density in Theta/Alpha Frequencies of the Waking EEG Progressively Increases During Sustained Wakefulness. *Sleep*, *18*(10), 890–894.

Caldwell, J. a. (2001). The impact of fatigue in air medical and other types of operations: a review of fatigue facts and potential countermeasures. *Air Medical Journal*, *20*(1), 25–32. https://doi.org/10.1067/mmj.2001.112420

Caldwell, J. A. (2005). Fatigue in aviation. *Travel Medicine and Infectious Disease*, *3*, 85–96. https://doi.org/10.1016/j.tmaid.2004.07.008

Caldwell, J. A. (2012). Crew Schedules, Sleep Deprivation, and Aviation Performance. *Current Directions in Psychological Science*, *21*(2), 85–89. https://doi.org/10.1177/0963721411435842

Caldwell, J. A., Caldwell, J. L., Smith, J. K., & Brown, D. L. (2004). Modafinil's effects on simulator performance and mood in pilots during 37 h without sleep. *Aviation Space and Environmental Medicine*, *75*(9), 777–784.

Caldwell, J. A., Caldwell, J. L., Thompson, L. A., & Lieberman, H. R. (2019). Fatigue and its management in the workplace. *Neuroscience & Biobehavioral Reviews*, *96*, 272–289. https://doi.org/10.1016/j.neubiorev.2018.10.024

Caldwell, J. A., Mallis, M. M., Caldwell, J. L., Paul, M. A., Miller, J. C., & Neri, D. F. (2009). Fatigue countermeasures in aviation. *Aviation Space and Environmental Medicine*, *80*, 29–59. https://doi.org/10.3357/ASEM.2435.2009

Caldwell, J. A., Mu, Q., Smith, J. K., Mishory, A., Caldwell, J. L., Peters, G., Brown, D. L., & George, M. S. (2005). Are individual differences in fatigue vulnerability related to baseline differences in cortical activation? *Behavioral Neuroscience, 119*(3), 694–707. https://doi.org/10.1037/0735-7044.119.3.694

Caldwell J, & Caldwell J. L. (2003). *Fatigue in Aviation: A Guide to Staying Awake at the Stick.* (1st ed.). Routledge.

Caldwell, J. A., Caldwell, J. L., Brown, D. L., & Smith, J. K. (2004). The Effects of 37 Hours of Continuous Wakefulness On the Physiological Arousal, Cognitive Performance, Self-Reported Mood, and Simulator Flight Performance of F-117A Pilots. *Military Psychology, 16*(3), 163–181. https://doi.org/10.1207/s15327876mp1603_2

Cao, T., Lian, Z., Ma, S., & Bao, J. (2021). Thermal comfort and sleep quality under temperature, relative humidity and illuminance in sleep environment. *Journal of Building Engineering*, *43*. https://doi.org/10.1016/j.jobe.2021.102575

Carskadon, M. A., & Dement, W. C. (2011). Normal Human Sleep: An Overview. In *Principles and practice of sleep medicine* (5th ed., pp. 16–26). Elseview Saunders.

Chang, A. M., Duffy, J. F., Buxton, O. M., Lane, J. M., Aeschbach, D., Anderson, C., Bjonnes, A. C., Cain, S. W., Cohen, D. A., Frayling, T. M., Gooley, J. J., Jones, S. E., Klerman, E. B., Lockley, S. W., Munch, M., Rajaratnam, S. M. W., Rueger, M., Rutter, M. K., Santhi, N., … Saxena, R. (2019). Chronotype Genetic Variant in PER2 is Associated with Intrinsic Circadian Period in Humans. *Nature: Scientific Reports*, *9*(1), 5350. https://doi.org/10.1038/s41598-019-41712-1

Chaput, J. P., Dutil, C., Featherstone, R., Ross, R., Giangregorio, L., Saunders, T. J., Janssen, I., Poitras, V. J., Kho, M. E., Ross-White, A., & Carrier, J. (2020). Sleep duration and health in adults: an overview of systematic reviews. *Applied Physiology, Nutrition, and Metabolism, 45*(10), 218–231. https://doi.org/10.1139/apnm-2020-0034

Chee, M. W. L., & Chuah, L. Y. M. (2008). Functional neuroimaging insights into how sleep and sleep deprivation affect memory and cognition. *Current Opinion in Neurology, 21*, 417–423.

Chee, M. W. L., Chuah, L. Y. M., Venkatraman, V., Chan, W. Y., Philip, P., & Dinges, D. F. (2006). Functional imaging of working memory following normal sleep and after 24 and

35 h of sleep deprivation: Correlations of fronto-parietal activation with performance. *NeuroImage*, *31*(1), 419–428. https://doi.org/10.1016/j.neuroimage.2005.12.001

Chee, M. W. L., Jiat, C. T., Zheng, H., Parimal, S., Weissman, D. H., Zagorodnov, V., & Dinges, D. F. (2008). Lapsing during sleep deprivation is associated with distributed changes in brain activation. *Journal of Neuroscience*, *28*(21), 5519–5528. https://doi.org/10.1523/JNEUROSCI.0733-08.2008

Chen, L., Zhao, Y., Zhang, J., & Zou, J. (2014). Investigation of Mental Fatigue Induced by a Continuous Mental Arithmetic Task Based on EEG Coherence Analysis. In S. Ma, L. Jia, X. Li, L. Wang, H. Zhou, & X. Sun (Eds.), *Life System Modeling and Simulation.* Springer.

Cheng, P., & Drake, C. (2016). Occupational Sleep Medicine. *Sleep Medicine Clinics*, *11*(1), 65–79. https://doi.org/10.1016/j.jsmc.2015.10.006

Choo, W.-C., Lee, W.-W., Venkatraman, V., Sheu, F.-S., & Chee, M. W. L. (2005). Dissociation of cortical regions modulated by both working memory load and sleep deprivation and by sleep deprivation alone. *NeuroImage*, *25*(2), 579–587. https://doi.org/10.1016/j.neuroimage.2004.11.029

Chuah, L. Y. M., & Chee, M. W. L. (2008). Cholinergic augmentation modulates visual task performance in sleep-deprived young adults. *Journal of Neuroscience*, *28*(44), 11369–11377. https://doi.org/10.1523/JNEUROSCI.4045-08.2008

Civil Aviation Authority. (2003a). *CAA Paper 2003/14. Wakefulness on the Civil Flight Deck: Evaluation of a Wrist-worn Alertness Device*. Retrieved 29/09/2016: http://publicapps.caa.co.uk/docs/33/CAPAP2003_14.PDF

Civil Aviation Authority. (2003b). *Paper 2003/8: A review of in-flight napping strategies* (Issue CAA Paper 2003/8). Retrieved 29/05/2018: https://publicapps.caa.co.uk/docs/33/2003_08.PDF

Civil Aviation Authority. (2016). *CAP 1265 - EASA Flight Time Limitations: GM1 CS FTL.1.225(b)(2)*. Retrieved 15/02/2019: http://publicapps.caa.co.uk/docs/33/CAP 1265 OCT16.pdf

Civil Aviation Authority. (2020). *Pilot Performance: Safety Plan*. Retrieved 21/08/2021: https://www.caa.co.uk/safety-initiatives-and-resources/how-we-regulate/safety-plan/mitigating-key-safety-risks/pilot-performance/.

Civil Aviation Authority. (2021a). *Respiratory System guidance material (MED.B. 015).* Retrieved 30/04/2022: https://www.caa.co.uk/aeromedical-examiners/medical-standards/pilots/conditions/respiratory/respiratory-guidance-material-gm/.

Civil Aviation Authority. (2021b). *Safety Notice: SN-2021/003 Fatigue Management - Guidance for AOC Holders During Covid-19 Changing Conditions*. Retrieved 30/04/2022: http://publicapps.caa.co.uk/docs/33/SN2021003%20Fatigue%20Management.pdf

Civil Aviation Authority (CAA). (2005). *Paper 2005/04. Aircrew Fatigue: A Review of Research Undertaken on Behalf of the UK Civil Aviation Authority*. Retrieved 05/11/2018: https://publicapps.caa.co.uk/docs/33/CAAPaper2005_04.pdf

Civil Aviation Authority (CAA). (2018). *Occurrence reporting: Guidance on mandatory and voluntary occurrence reports (MORs and VORs)*. Retrieved 19/12/18: https://www.caa.co.uk/our-work/make-a-report-or-complaint/mor/occurrence-reporting/.

Civil Aviation Authority (CAA). (2022). *CAA Ramp inspection programme: Alcohol testing on flight crew and cabin crew members in the United Kingdom*. Retrieved 06/04/22: https://www.caa.co.uk/commercial-industry/airlines/uk-ramp-inspection-programme/.

Civil Aviation Safety Authority Australia. (2014). *Biomathematical Fatigue Model Guidance Document*. Retrieved 18/06/2020: http://www.test.casa.gov.au/wcmswr/_assets/main/aoc/fatigue/fatigue_modelling.pdf

Colman, A. M. (2006). *Oxford Dictionary of Psychology: Definition of fatigue* (2nd ed.). Oxford University Press.

ComRes. (2013). *ComRes Report prepared on behalf of the British Airline Pilots' Association: BALPA 2013 Membership Survey*. Retrieved 07/01/2018: https://www.comresglobal.com/wp-content/themes/comres/poll/BALPA_Membership_Survey.pdf

ComRes. (2016). *Aeromedical Examiner Research*. Retrieved 15/12/2016: https://comresglobal.com/

Conklin, T., 2012. *Pre-accident investigations: An introduction to organizational safety*. Ashgate Publishing Ltd.

Cooke, D. L., & Rohleder, T. R. (2006). Learning from incidents: from normal accidents to high reliability. *System Dynamics Review, 22*(3), 213–239.

Corsi-Cabrera, M., Sanchez, A., del Rio Portilla, Y., & Villanueva, Y. (2003). Effect of 38 h of total sleep deprivation on the waking EEG in women: sex differences. *International Journal of Psychophysiology*, *50*(3), 213–224. https://doi.org/10.1016/S0167-8760(03)00168-5

Costa, G. (2015). Sleep deprivation due to shift work. In *Handbook of Clinical Neurology* (1st ed., Vol. 131). Elsevier B.V. https://doi.org/10.1016/B978-0-444-62627-1.00023-8

Court of Inquiry India. (2010). *Accident to Air India Express Boeing 737-800 aircraft VT-AXV on 22nd May 2010 at Mangalore*. Retrieved 18/05/2022 https://www.skybrary.aero/sites/default/files/bookshelf/1680.pdf

Curcio, G., Casagrande, M., & Bertini, M. (2001). Sleepiness: evaluating and quantifying methods. *International Journal of Psychophysiology*, *41*, 251–263.

Curcio, G., Ferrara, M., & de Gennaro, L. (2006). Sleep loss, learning capacity and academic performance. *Sleep Medicine Reviews*, *10*(5), 323–337. https://doi.org/10.1016/j.smrv.2005.11.001

Czeisler, C. A., Duffy, J. F., Shanahan, T. L., Brown, E. N., Mitchell, J. F., Rimmer, D. W., Ronda, J. M., Silva, E. J., Allan, J. S., Emens, J. S., Dijk, D.-J., & Kronauer, R. E. (1999). Stability, Precision, and Near-24-Hour Period of the Human Circadian Pacemaker. *Science*, *284*(5423), 2177–2181. https://doi.org/10.1126/science.284.5423.2177

Czeisler, C. A., Weitzman, E. D., Moore-Ede, M. C., Zimmerman, J. C., & Knauer, R. S. (1980). Human Sleep: Its Duration and Organization Depend on Its Circadian Phase. *Science*, *210*(4475), 1264–1267. https://doi.org/10.1126/science.7434029

Dahlgren, K. (1981). Adjustment of Circadian Rhythms and EEG Sleep Functions to Day and Night Sleep Among Permanent Nightworkers and Rotating Shiftworkers. *Psychophysiology*, *18*(4), 381–391. https://doi.org/10.1111/j.1469-8986.1981.tb02469.x

Darwent, D., Dawson, D., Paterson, J. L., Roach, G. D., & Ferguson, S. A. (2015). Managing fatigue: It really is about sleep. *Accident Analysis and Prevention*, *82,* 20–26. https://doi.org/10.1016/j.aap.2015.05.009

Darwent, D., Dawson, D., & Roach, G. D. (2010). Prediction of Probabilistic Sleep Distributions Following Travel Across Multiple Time Zones Predicting Sleep After Multiple Time Zone Crossings. *SLEEP*, *33*(2), 185–195. https://academic.oup.com/sleep/article/33/2/185/2454484

Darwent, D., Dawson, D., & Roach, G. D. (2012). A model of shiftworker sleep/wake behaviour. *Accident Analysis and Prevention*, *45*(SUPPL.), 6–10. https://doi.org/10.1016/j.aap.2011.09.017

Dawson, D. (2012). Fatigue research in 2011: From the bench to practice. In *Accident Analysis and Prevention*, *45*, 1–5. https://doi.org/10.1016/j.aap.2011.09.016

Dawson, D., Chapman, J., & Thomas, M. J. W. (2012). Fatigue-proofing: A new approach to reducing fatigue-related risk using the principles of error management. *Sleep Medicine Reviews*, *16*, 167–175. https://doi.org/10.1016/j.smrv.2011.05.004

Dawson, D., Cleggett, C., Thompson, K., & Thomas, M. J. W. (2017). Fatigue proofing: The role of protective behaviours in mediating fatigue-related risk in a defence aviation environment. *Accident Analysis & Prevention*, *99*, 465–468. https://doi.org/10.1016/j.aap.2015.10.011

Dawson, D., Ian Noy, Y., Härmä, M., Åkerstedt, T., & Belenky, G. (2011). Modelling fatigue and the use of fatigue models in work settings. *Accident Analysis & Prevention*, *43*(2), 549–564. https://doi.org/10.1016/j.aap.2009.12.030

Dawson, D., & McCulloch, K. (2005). Managing fatigue: It's about sleep. *Sleep Medicine Reviews*, *9*, 365–380. https://doi.org/10.1016/j.smrv.2005.03.002

Dawson, D., & Reid, K. (1997). Fatigue, alcohol and performance impairment. *Nature*, *388*(6639), 235–235. https://doi.org/10.1038/40775

de Gennaro, L., Ferrara, M., Ferlazzo, F., & Bertini, M. (2000). Slow eye movements and EEG power spectra during wake-sleep transition. *Clinical Neurophysiology*, *111*(12), 2107–2115. https://doi.org/10.1016/S1388-2457(00)00476-4

Dejohn, C. A., Wolbrink, A. M., & Larcher, J. G. (2004). *FAA Report: In-Flight Medical Incapacitation and Impairment of U.S. Airline Pilots: 1993 to 1998* (Issue October). Retrieved: https://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2000s/media/0416.pdf

Dekker, S. (2002). *The Field Guide to Human Error Investigations*. Ashgate.

Dekker, S. (2015). *Safety Differently: Human Factors for a New Era*. (2nd ed). Taylor & Francis Group.

della Rocco, P. S. (1999). The Role of Shift Work and Fatigue in Air Traffic Control Operational Errors and Incidents. *Federal Aviation Administration: Office of Aviation Medicine*.

Department of Transportation. (2000). *Hours of Service of Drivers; Driver Rest and Sleep for Safe Operations; Proposed Rule*. https://www.federalregister.gov/documents/2000/05/02/00-10703/hours-of-service-of-drivers-driver-rest-and-sleep-for-safe-operations

Devine, J. K., Choynowski, J., Garcia, C. R., Simoes, A. S., Guelere, M. R., Godoy, B. de, Silva, D. S., Pacheco, P., & Hursh, S. R. (2021). Pilot Sleep Behavior across Time during Ultra-Long-Range Flights. *Clocks & Sleep*, *3*(4), 515–527. https://doi.org/10.3390/clockssleep3040036

Dijk, D. J., & Archer, S. N. (2009). Light, sleep, and circadian rhythms: Together again. In *PLoS Biology* (Vol. 7, Issue 6). https://doi.org/10.1371/journal.pbio.1000145

Dinges, D. F. (1986). Differential effects of prior wakefulness and circadian phase on nap sleep. *Electroencephalography and Clinical Neurophysiology, 64*, 224–227.

Dinges, D. F. (1990a). Are you awake? Cognitive performance and reverie during the hypnopompic state. In R. R. Bootzin, J. F. Kihlstrom, & D. L. Schacter (Eds.), *Sleep and Cognition* (pp. 159–175). American Psychological Association.

Dinges, D. F. (1990b). The nature of subtle fatigue effects in long-haul crews. *Proceedings of the Flight Safety Foundation 43rd International Air Safety Seminar*, 258–267.

Dinges, D. F. (1995). An Overview of Sleepiness and Accidents. *Journal of Sleep Research*, *4*(Suppl.2), 4–14.

Dinges, D. F. (2004). Critical Research Issues in Development of Biomathematical Models of Fatigue and Performance. *Aviation Space and Environmental Medicine*, *75*(3), 181–191.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., Aptowicz, C., & Pack, A. I. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep*, *20*(4), 267–277.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers*, *17*(6), 652–655. https://doi.org/10.3758/BF03200977

Dittner, A. J., Wessely, S. C., & Brown, R. G. (2004). The assessment of fatigue: A practical guide for clinicians and researchers. *Journal of Psychosomatic Research*, *56*(2), 157–170. https://doi.org/10.1016/S0022-3999(03)00371-4

Doran, S. M., van Dongen, H. P. , & Dinges, D. F. (2001). Sustained attention performance during sleep deprivation: Evidence of state instability. . *Archives Italiennes de Biologie* , *139*, 253–267.

Dörr, J., & Nater, U. (2013). Erschöpfungssyndrome - Eine Diskussion verschiedener Begriffe, Definitionsansätze und klassifikatorischer Konzepte. *PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie*, *63*(02), 69–76. https://doi.org/10.1055/s-0032-1327706

Dorrian, J., Darwent, D., Dawson, D., & Roach, G. D. (2012). Predicting pilot's sleep during layovers using their own behaviour or data from colleagues: Implications for biomathematical models. *Accident Analysis and Prevention*, *45*(SUPPL.), 17–21. https://doi.org/10.1016/j.aap.2011.09.019

Dorrian, J., Rogers, N. L., & Dinges, D. F. (2005). Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss. In *Sleep Deprivation* (Kushida, C.A., pp. 39–70). Marcel Dekker.

Doudou, M., Bouabdallah, A., & Berge-Cherfaoui, V. (2020). Driver Drowsiness Measurement Technologies: Current Research, Market Solutions, and Challenges. *International Journal of Intelligent Transportation Systems Research*, *18*(2), 297–319. https://doi.org/10.1007/s13177-019-00199-w

Downer, J. (2009). *When Failure is an Option: Redundancy, reliability and regulation in complex technical systems.* CARR Discussion Papers (DP 53). Centre for Analysis of Risk and Regulation: London School of Economics and Political Science.

Drummond, S. P. A., Anderson, D. E., Straus, L. D., Vogel, E. K., & Perez, V. B. (2012). The Effects of Two Types of Sleep Deprivation on Visual Working Memory Capacity and Filtering Efficiency. *PLoS ONE*, *7*(4), e35653. https://doi.org/10.1371/journal.pone.0035653

Drummond, S. P. A., Brown, G. G., Gillin, J. C., Stricker, J. L., Wong, E. C., & Buxton, R. B. (2000). Altered brain response to verbal learning following sleep deprivation. *Nature*, *403*(6770), 655–657. https://doi.org/10.1038/35001068

Drummond, S. P. A., Brown, G. G., Salamat, J. S., & Gillin, J. C. (2004). Increasing task difficulty facilitates the cerebral compensatory response to total sleep deprivation. *Sleep*, *27*(3), 445–451.

Drummond, S. P. A., Meloy, M. J., Yanagi, M. A., Orff, H. J., & Brown, G. G. (2005). Compensatory recruitment after sleep deprivation and the relationship with performance. *Psychiatry Research: Neuroimaging*, *140*(3), 211–223. https://doi.org/10.1016/j.pscychresns.2005.06.007

Drummond S, Paulus M, & Tapert S. (2006). Effects of two nights sleep deprivation and two nights recovery sleep on response inhibition. *Journal of Sleep Research*, *15*(3), 261–265. https://doi.org/10.1111/j.1365-2869.2006.00535.x

Drury, D. A., Ferguson, S. A., & Thomas, M. J. W. (2012). Restricted sleep and negative affective states in commercial pilots during short haul operations. *Accident Analysis and Prevention*, *45*(SUPPL.), 80–84. https://doi.org/10.1016/j.aap.2011.09.031

Duffy, J. F., & Czeisler, C. A. (2002). Age-related change in the relationship between circadian period, circadian phase, and diurnal preference in humans. *Neuroscience Letters*, *318*(3), 117–120.

Durmer, J. S., & Dinges, D. F. (2005). Neurocognitive Consequences of Sleep Deprivation. *Seminars in Neurology*, *25*(1), 117–129. http://faculty.vet.upenn.edu/uep/user_documents/dfd3.pdf

Ekstedt, M., Soderstrom, M., & Åkerstedt, T. (2009). Sleep physiology in recovery from burnout. *Biological Psychology*, *82*(3), 267–273.

Elmenhorst, E. M., Elmenhorst, D., Luks, N., Maass, H., Vejvoda, M., & Samel, A. (2008). Partial sleep deprivation: Impact on the architecture and quality of sleep. *Sleep Medicine*, *9*(8), 840–850. https://doi.org/10.1016/j.sleep.2007.07.021

Endsley, M. R., Farley, T. C., Jones, W. M., Midkiff, A. H., & Hansman, R. J. (1998). *Situation awareness information requirements for commercial airline pilots*. International Center for Air Transportation: MIT Press

Eriksen, C. A., & Åkerstedt, T. (2006). Aircrew fatigue in trans-Atlantic morning and evening flights. *Chronobiology International*, *23*(4), 843–858. https://doi.org/10.1080/07420520600865574

Eskandarian, A., & Mortazavi, A. (2007). Evaluation of a Smart Algorithm for Commercial Vehicle Driver Drowsiness Detection. *2007 IEEE Intelligent Vehicles Symposium*, 553–559. https://doi.org/10.1109/IVS.2007.4290173

European Aviation Safety Agency, E. (2014). 1st Workshop on the Approval and Oversight of Fatigue Risk Management (FRM). *The Approval and Oversight of Fatigue Risk Management (FRM)*. https://www.easa.europa.eu/newsroom-and-events/events/1st-workshop-approval-and-oversight-fatigue-risk-management-frm

European Cockpit Association. (2012). *ECA fatigue survey barometer*. https://www.eurocockpit.be/sites/default/files/eca_barometer_on_pilot_fatigue_12_1107_f.pdf

Commission Regulation (EU) No 83/2014, Official Journal of the European Union (2014).

Commission implementing regulation (EU) 2015/2018, Official Journal of the European Union (2015).

Evans, S., & Radcliffe, S.-A. (2012). The Annual Incapacitation Rate of Commercial Pilots. *Aviation, Space, and Environmental Medicine*, *83*(1), 42–49. https://doi.org/10.3357/ASEM.3134.2012

Facer-Childs, E. R., Boiling, S., & Balanos, G. M. (2018). The effects of time of day and chronotype on cognitive and physical performance in healthy volunteers. *Sports Medicine*, *4*(47).

Fard, S. J., & Lavender, A. P. (2019). A comparison of task-based mental fatigue between healthy males and females. *Fatigue: Biomedicine, Health and Behaviour*, *7*(1), 1–11.

Flight crew member duty and rest requirements, US DOT FAA Docket No. FAA-2009-1093 121 (2011).

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications Ltd.

Fink, A., Grabner, R. H., Benedek, M., Reishofer, G., Hauswirth, V., Fally, M., Neuper, C., Ebner, F., & Neubauer, A. C. (2009). The creative brain: Investigation of brain activity during creative problem solving by means of EEG and FMRI. *Human Brain Mapping*, *30*(3), 734–748. https://doi.org/10.1002/hbm.20538

Flynn-Evans, E. E., Arsintescu, L., Gregory, K., Mulligan, J., Nowinski, J., & Feary, M. (2018). Sleep and neurobehavioral performance vary by work start time during non-traditional day shifts. *Sleep Health*, *4*(5), 476–484. https://doi.org/10.1016/j.sleh.2018.08.002

Folkard, S. (2008). Do Permanent Night Workers Show Circadian Adjustment? A Review Based on the Endogenous Melatonin Rhythm. *Chronobiology International*, *25*(2–3), 215–224. https://doi.org/10.1080/07420520802106835

Folkard, S., & Åkerstedt, T. (2004). Trends in the risk of accidents and injuries and their implications for models of fatigue and performance. *Aviation Space and Environmental Medicine*, *75*.

Folkard, S., & Barton, J. (1993). Does the “forbidden zone” for sleep onset influence morning shift sleep duration? *Ergonomics*, *36*, 85–91.

Folkard, S., Lombardi, D. A., & Spencer, M. B. (2006). Estimating the Circadian Rhythm in the Risk of Occupational Injuries and Accidents. *Chronobiology International*, *23*(6), 1181–1192. https://doi.org/10.1080/07420520601096443

Folkard, S., Lombardi, D. A., & Tucker, P. T. (2005). Shiftwork: Safety, Sleepiness and Sleep. *Industrial Health*, *43*(1), 20–23. https://doi.org/10.2486/indhealth.43.20

Frey, D. J., Badia, P., & Wright, K. P. (2004). Inter- and intra-individual variability in performance near the circadian nadir during sleep deprivation. *Journal of Sleep Research*, *13*(4), 305–315. https://doi.org/10.1111/j.1365-2869.2004.00429.x

Ftouni, S., Rahman, S. A., Crowley, K. E., Anderson, C., Rajaratnam, S. M. W., & Lockley, S. W. (2013). Temporal Dynamics of Ocular Indicators of Sleepiness across Sleep Restriction. *Journal of Biological Rhythms*, *28*(6), 412–424. https://doi.org/10.1177/0748730413512257

Galland-Decker, C., Marques-Vidal, P., & Vollenweider, P. (2019). Prevalence and factors associated with fatigue in the Lausanne middle-aged population: a population-based, cross-sectional survey. *BMJ Open*, *9*(8), e027070. https://doi.org/10.1136/bmjopen-2018-027070

Gander P, Graeber R, Connell L, Gregory K, Miller D, & Rosekind M. (1998). Flight crew fatigue II: short-haul fixed-wing air transport operations. *Aviation, Space and Environmental Medicine*, *69*(Suppl 9), B8-15.

Gander, P. H. (2015). Evolving Regulatory Approaches for Managing Fatigue Risk in Transport Operations. *Reviews of Human Factors and Ergonomics*, *10*(1), 253–271. https://doi.org/10.1177/1557234X15576510

Gander, P. H., Mulrine, H. M., van den Berg, M. J., Smith, A. A. T., Signal, T. L., Wu, L. J., & Belenky, G. (2014). Pilot fatigue: Relationships with departure and arrival times, flight duration, and direction. *Aviation Space and Environmental Medicine, 85*(8), 833–840. https://doi.org/10.3357/ASEM.3963.2014

Gander, P. H., Mulrine, H. M., van den Berg, M. J., Smith, A. A. T., Signal, T. L., Wu, L. J., & Belenky, G. (2015). Effects of sleep/wake history and circadian phase on proposed pilot fatigue safety performance indicators. *Journal of Sleep Research, 24*(1), 110–119. https://doi.org/10.1111/jsr.12197

Gander, P. H., Nguyen, D., Rosekind, M. R., & Connell, L. J. (1993). Age, circadian rhythms, and sleep loss in flight crews. *Aviation, Space, and Environmental Medicine, 64*(3), 189195.

Gander, P. H., Rosekind, M. R., & Gregory, K. B. (1998). Flight crew fatigue VI: a synthesis. *Aviation, Space, and Environmental Medicine, 69*(9), 49–60.

Gander, P. H., Signal, T. L., van den Berg, M. J., Mulrine, H. M., Jay, S. M., & Jim Mangie, C. (2013a). In-flight sleep, pilot fatigue and Psychomotor Vigilance Task performance on ultra-long range versus long range flights. *Journal of Sleep Research*, *22*(6), 697–706. https://doi.org/10.1111/jsr.12071

Gander, P., Hartley, L., Powell, D., Cabon, P., Hitchcock, E., Mills, A., & Popkin, S. (2011). Fatigue risk management: Organizational factors at the regulatory and industry/company level. *Accident Analysis and Prevention*, *43,* 573–590. https://doi.org/10.1016/j.aap.2009.11.007

Gander, P., & Signal, L. (2008). Who is too old for shift work? Developing better criteria. *Chronobiology International*, *25*(2–3), 199–213. https://doi.org/10.1080/07420520802077556

Gander, P., van den Berg, M., Mulrine, H., Signal, L., & Mangie, J. (2013). Circadian adaptation of airline pilots during extended duration operations between the USA and Asia. *Chronobiology International*, *30*(8), 963–972. https://doi.org/10.3109/07420528.2013.790042

Gawron, V. J. (2016). Summary of Fatigue Research for Civilian and Military Pilots. *IIE Transactions on Occupational Ergonomics and Human Factors*, *4*(1), 1–18. https://doi.org/10.1080/21577323.2015.1046093

Gillberg, M., Kecklund, G., Goransson, B., & Åkerstedt, T. (2003). Operator performance and signs of sleepiness during day and night work in a simulated thermal power plant. *International Journal of Industrial Ergonomics*, *31*(2), 101-109.

Gimeno, P. T., Cerezuela, G. P., & Montanes, M. C. (2006). On the concept and measurement of driver drowsiness, fatigue and inattention: implications for countermeasures. *International Journal of Vehical Design*, *42*(1), 67–86.

Goel, N. (2017). Neurobehavioral Effects and Biomarkers of Sleep Loss in Healthy Adults. In *Current Neurology and Neuroscience Reports* (Vol. 17, Issue 11). Current Medicine Group LLC 1. https://doi.org/10.1007/s11910-017-0799-x

Goel, N., Basner, M., & Dinges, D. F. (2015). Phenotyping of neurobehavioral vulnerability to circadian phase during sleep loss. *Methods in Enzymology*, 552, 285–308. https://doi.org/10.1016/bs.mie.2014.10.024

Goel, N., Basner, M., Rao, H., & Dinges, D. F. (2013). Circadian rhythms, sleep deprivation, and human performance. *Progress in Molecular Biology and Translational Science*, *119*, 155-190. https://doi.org/10.1016/B978-0-12-396971-2.00007-5)

Goel, N., Rao, H., Durmer, J. S., & Dinges, D. F. (2009). Neurocognitive consequences of sleep deprivation. *Seminars in Neurology*, *29*(4), 320–339. https://doi.org/10.1055/s-0029-1237117

Goel, V., & Vartanian, O. (2005). Dissociating the roles of right ventral lateral and dorsal lateral prefrontal cortex in generation and maintenance of hypotheses in set-shift problems. *Cerebral Cortex*, *15*(8), 1170–1177. https://doi.org/10.1093/cercor/bhh217

Goode, J. H. (2003). Are pilots at risk of accidents due to fatigue? *Journal of Safety Research*, *34*, 309–313. https://doi.org/10.1016/S0022-4375(03)00033-1

Graeber, R. C. (2008). Fatigue risk management systems within SMS. In *Proceedings of the FAA Fatigue Management Symposium: Partnerships for Solutions* (pp. 17-19).

Graeber, R. C., Rosekind, M. R., Connell, L. J., & Dinges, D. (1990). Cockpit napping. *ICAO Journal*, *10,* 6–10.

Grant, E., Salmon, P. M., Stevens, N. J., Goode, N., & Read, G. J. (2018). Back to the future: What do accident causation models tell us about accident prediction? *Safety Science*, *104*, 99–109. https://doi.org/10.1016/j.ssci.2017.12.018

Greenberg, D. B. (2002). Clinical Dimensions of Fatigue. *Journal of Clinical Psychiatry*, *4*, 90–93.

Gregory, K. B., Winn, W., Johnson, K., & Rosekind, M. R. (2010). Pilot fatigue survey: Exploring fatigue factors in air medical operations. *Air Medical Journal*, *29*(6), 309–319. https://doi.org/10.1016/j.amj.2010.07.002

Griepentrog, J. E., Labiner, H. E., Gunn, S. R., & Rosengart, M. R. (2018). Bright environmental light improves the sleepiness of nightshift ICU nurses. *Critical Care*, *22*(1). https://doi.org/10.1186/s13054-018-2233-4

Groeger, J. A., Zijlstra, F. R. H., & Dijk, D. J. (2004). Sleep quantity, sleep difficulties and their perceived consequences in a representative sample of some 2000 British adults. *Journal of Sleep Research*, *13*, 359–371.

Gundel, A., Drescher, J., Maab, H., Samel, A., & Vejvoda, M. (1995). Sleepiness of civil airline pilots during two consecutive night flights of extended duration. In *Biological Psychology* (Vol. 40).

Gundel, A., Marsalek, K., & ten Thoren, C. (2007). A critical review of existing mathematical models for alertness. *Somnologie - Schlafforschung Und Schlafmedizin*, *11*(3), 148–156. https://doi.org/10.1007/s11818-007-0312-x

Haidarimoghadam, R., Kazemi, R., Motamedzadeh, M., Golmohamadi, R., Soltanian, A., & Zoghipaydar, M. R. (2017). The effects of consecutive night shifts and shift length on cognitive performance and sleepiness: a field study. *International Journal of Occupational Safety and Ergonomics*, *23*(2), 251–258. https://doi.org/10.1080/10803548.2016.1244422

Härmä, M., Sallinen, M., Ranta, R., Mutanen, P., & Müller, K. (2002). The effect of an irregular shift system on sleepiness at work in train drivers and railway traffic controllers. *Journal of Sleep Research, 11*(2), 141-151. https://doi.org/10.1046/j.1365-2869.2002.00294.x

Harris W, Sachau D, & Hanis S. (2001). The relationship between working conditions and commercial pilot fatigue development. *Proceedings of the Human Factors and Ergonomics Society*, 185–188.

Harrison, Y., & Horne, J. A. (1996). Occurrence of “microsleeps” during daytime sleep onset in normal subjects. *Electroencephalography and Clinical Neurophysiology*, *98*, 4–14.

Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: A review. *Journal of Experimental Psychology: Applied*, *6*(3), 236–249. https://doi.org/10.1037/1076-898X.6.3.236

Hartzler, B. M. (2014). Fatigue on the flight deck: The consequences of sleep loss and the benefits of napping. *Accident Analysis and Prevention*, *62*, 309–318. https://doi.org/10.1016/j.aap.2013.10.010

Hastings, M., O’Neill, J. S., & Maywood, E. S. (2007). Circadian clocks: Regulators of endocrine and metabolic rhythms. *Journal of Endocrinology*, *195*(2), 187–198. https://doi.org/10.1677/JOE-07-0378

Heuer, H., Kohlisch, O., & Klein, W. (2005). The Effects of Total Sleep Deprivation on the Generation of Random Sequences of Key-Presses, Numbers and Nouns. *The Quarterly Journal of Experimental Psychology Section A*, *58*(2), 275–307. https://doi.org/10.1080/02724980343000855

Hockey, R. (2012). *The Psychology of Fatigue: Work, Effort and Control*. Cambridge University Press.

Hoddes, E., Dement, W., & Zarcone, V. (1972). The development and use of the Stanford sleepiness scale (SSS). *Psychophysiology*, *9*(150).

Hollnagel, E. (2004). *Barriers and Accident Prevention* (E. Hollnagel, Ed.; 1st ed.). Routledge.

Hollnagel, E. (2018). *Safety-I and Safety-II: The Past and Future of Safety Management.* New York: CRC Press, Taylor and Francis Group.

Hollnagel, E., Wears, R.L. & Braithwaite, J. (2015). *From Safety I to Safety II: A White Paper.* The Resilient Health Care Net: Published simultaneously by the University of Southern Denmark, University of Florida, USA, and Macquarie University, Australia.

Honkus, V. L. (2003). Sleep Deprivation in Critical Care Units. *Critical Care Nursing Quarterly*, *26*(3), 179–191.

Honn, K. A., Satterfield, B. C., McCauley, P., Caldwell, J. L., & van Dongen, H. P. A. (2016). Fatiguing effect of multiple take-offs and landings in regional airline operations. *Accident Analysis and Prevention, 86*, 199–208. https://doi.org/10.1016/j.aap.2015.10.005

Horne, J. (1988). *Why we sleep: The functions of sleep in humans and other mammals*. Oxford University Press.

Horne, J. A. (1988). Sleep Loss and “Divergent” Thinking Ability. *Sleep, 11*(6), 528–536. https://doi.org/10.1093/sleep/11.6.528

Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, *4*(2), 97–110.

Horne, J. A., & Pettitt, A. N. (1985). High incentive effects on vigilance performance during 72 hours of total sleep deprivation. *Acta Psychologica*, *58*(2), 123–139. https://doi.org/10.1016/0001-6918(85)90003-4

Horne, J. A., & Reyner, L. A. (1995a). Sleep related vehicle accidents. *BMJ*, *310*(6979), 565–567. https://doi.org/10.1136/bmj.310.6979.565

Horne, J. A., & Reyner, L. A. (1995b). Driver sleepiness. *Journal of Sleep Research*, *4*, 23–29. https://doi.org/10.1111/j.1365-2869.1995.tb00222.x

Horne, J., & Reyner, L. (1999). Vehicle accidents related to sleep: A review. *Occupational and Environmental Medicine*, *56*(5), 289–294. https://doi.org/10.1136/oem.56.5.289

Houston, S., Dawson, K., & Butler, S. (2012). Fatigue reporting among aircrew: Incidence rate and primary causes. *Aviation Space and Environmental Medicine*, *83*(8), 800–804. https://doi.org/10.3357/ASEM.3238.2012

Huang, Z.-L., Urade, Y., & Hayaishi, O. (2011). The Role of Adenosine in the Regulation of Sleep. *Current Topics in Medicinal Chemistry, 11*(8), 1047–1057. https://doi.org/10.2174/156802611795347654

Hunter, R. (2015). Illusions of safety. In *Risk and Regulation: LSE Centre for Analysis of Risk and Regulation (CARR).*

Hursh, S. R., & van Dongen, H. P. A. (2011). Fatigue and Performance Modelling. In *Principles and Practice of Sleep Medicine* (pp. 745–752). Elsevier. https://doi.org/10.1016/B978-1-4160-6645-3.00066-9

IATA (2014). *Fatigue Safety Performance Indicators (SPIs): A Key Component of Proactive Fatigue Hazard Identification.* Retrieved 11/09/16:

https://www.iata.org/contentassets/5f976bb3ca2446f3a40e88b18dd61fbb/fatigue-spis_a-key-component-of-proactive-fatigue-hazard-identification.pdf

IATA. (2020). *Guidance for managing crew fatigue during a crisis*. Retrieved 01/07/2022: https://www.iata.org/contentassets/d0e499e4b2824d4d867a8e07800b14bd/iata-guidance-managing-crew-fatigue-during-crisis.pdf

IATA, ICAO & IFALPA. (2011). *Fatigue Risk Management Systems: Implementation Guide for Operators.* Retrieved 01/07/16: https://www.icao.int/safety/fatiguemanagement/frms%20tools/frms%20implementation%20guide%20for%20operators%20july%202011.pdf

ICAO..(2012).*Safety Management Manual (SMM): Doc 9859*. (3rd Edition). Retrieved 22/07/16: https://www.icao.int/sam/documents/rst-smsssp-13/smm_3rd_ed_advance.pdf

ICAO. (2020). *Manual for the Oversight of Fatigue Management Approaches - Doc 9966.* Retrieved 14/12/2020: https://www.icao.int/safety/fatiguemanagement/FRMS%20Tools/Doc%209966.FRMS.2016%20Edition.en.pdf

Ingre, M., Åkerstedt, T., Peters, B., Anund, A., & Kecklund, G. (2006). Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences. *Journal of Sleep Research*, *15*(1), 47-53. https://doi.org/10.1111/j.1365-2869.2006.00504.x

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. https://doi.org/10.1111/j.1365-2869.2006.00517.x

Ingre, M., Kecklund, G., Åkerstedt, T., & Kecklund, L. (2004). Variation in sleepiness during early morning shifts: a mixed model approach to an experimental field study of train drivers. *Chronobiology International*, *21*, 973–990.

Ingre, M., Kecklund, G., Åkerstedt, T., Söderström, M., & Kecklund, L. (2008). Sleep length as a function of morning shift-start time in irregular shift schedules for train drivers: Self-rated health and individual differences. *Chronobiology International*, *25*(2–3), 349–358. https://doi.org/10.1080/07420520802110704

Ingre, M., van Leeuwen, W., Klemets, T., Ullvetter, C., Hough, S., Kecklund, G., Karlsson, D., & Åkerstedt, T. (2014). Validating and extending the three process model of alertness in airline operations. *PLoS ONE*, *9*(10). https://doi.org/10.1371/journal.pone.0108679

International Civil Aviation Organisation, & ICAO. (2012). *Manual of Civil Aviation Medicine: Doc 8983 AN/895* (3rd ed.).

Jackson, C. A., & Earle, L. (2006). Prevalence of fatigue among commercial pilots. *Occupational Medicine*, *56*(4), 263–268. https://doi.org/10.1093/occmed/kql021

Jackson, M. L., Kennedy, G. A., Clarke, C., Gullo, M., Swann, P., Downey, L. A., Hayley, A. C., Pierce, R. J., & Howard, M. E. (2016). The utility of automated measures of ocular metrics for detecting driver drowsiness during extended wakefulness. *Accident Analysis and Prevention*, *87*, 127–133. https://doi.org/10.1016/j.aap.2015.11.033

Jackson, M. L., Raj, S., Croft, R. J., Hayley, A. C., & Downey, L. (2016). Slow eyelid closure as a measure of driver drowsiness and its relationship to performance Publication Details. *Traffic Injury Prevention*, *17*(3), 251–257. http://ro.uow.edu.au/sspapers/2207

Jennings, J. R., Monk, T. H., & van der Molen, M. W. (2003). Sleep Deprivation Influences Some but Not All Processes of Supervisory Attention. *Psychological Science*, *14*(5), 473–486. https://doi.org/10.1111/1467-9280.02456

Johns, M. W. (1991). A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep*, *14*(6), 540-545. https://doi.org/10.1093/sleep/14.6.540

Johns, M. W., Chapman, R., Crowley, K., & Tucker, A. (2008). A new method for assessing the risks of drowsiness while driving. *Somnologie*, *12*(1), 66–74.

Johns, M. W., Tucker, A., Chapman, R., Crowley, K., & Michael, N. (2007). Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers. *Somnologie*, *11*(4), 234–242. https://doi.org/10.1007/s11818-007-0311-y

Jonelis, M. B., Drummond, S. P. A., Salamat, J. S., McKenna, B. S., Ancoli-Israel, S., & Bondi, M. W. (2012). Age-Related Influences of Prior Sleep on Brain Activation during Verbal Encoding. *Frontiers in Neurology*, *3*. https://doi.org/10.3389/fneur.2012.00049

Juda, M., Vetter, C., & Roenneberg, T. (2013). The Munich ChronoType Questionnaire for shift-workers (MCTQ Shift). *Journal of Biological Rhythms*, *28*(2), 130–140. https://doi.org/10.1177/0748730412475041

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006a). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, *117*(7), 1574–1581. https://doi.org/10.1016/j.clinph.2006.03.011

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006b). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2006.03.011

Kandelaars, K., Dorrian, J., Fletcher, A., Roach, G. D., & Dawson, D. (2005). A review of biomathematical fatigue models: Where to from here? *Sixth International Conference on Fatigue Management in Transportation*, 3–16.

Kecklund, G., & Åkerstedt, T. (1993). Sleepiness in long distance truck driving: an ambulatory EEG study of night driving. *Ergonomics*, *36*(9), 1007–1017. https://doi.org/10.1080/00140139308967973

Kecklund, G., & Åkerstedt, T. (1995). Effects of timing of shifts on sleepiness and sleep duration. *Journal of Sleep Research*, *4*(Suppl. 2), 47–50.

Kecklund, G., Åkerstedt, T., & Lowden, A. (1997). Morning Work: Effects of Early Rising on Sleep and Alertness. *Sleep*, *20*(3), 215–223.

Kecklund, G., Gillberg, M., & Åkerstedt, T. (1994). Accumulated time with sleepiness: a new instrument for quantification of subjective sleepiness. *Journal of Sleep Research*, *3*(Suppl 1), 125.

Kendall A, Kautz M, Russo M, & Killgore W. (2006). Effects of sleep deprivation on lateral visual attention. *International Journal of Neuroscience*, *116*(10), 1125–1138. https://doi.org/10.1080/00207450500513922

Kerkhof, G. A., & van Dongen, H. P. A. (1996). Morning-type and evening-type individuals differ in the phase position of their endogenous circadian oscillator. *Neuroscience Letters*, *218*, 153–156.

Kern A. (1997). *Redefining airmanship*. McGraw-Hill.

Killgore, W. D. S. (2010). Effects of sleep deprivation on cognition. *Progress in Brain Research*, *185*, 105–129. https://doi.org/10.1016/B978-0-444-53702-7.00007-5

Killgore, W. D. S., Kahn-Greene, E. T., Grugle, N. L., Killgore, D. B., & Balkin, T. J. (2009). Sustaining Executive Functions During Sleep Deprivation: A Comparison of Caffeine, Dextroamphetamine, and Modafinil. *Sleep*, *32*(2), 205–216. https://doi.org/10.1093/sleep/32.2.205

Kim, Y., Elmenhorst, D., Weisshaupt, A., Wedekind, F., Kroll, T., Mccarley, R. W., Strecker, R. E., & Bauer, A. (2015). Chronic sleep restriction induces long-lasting changes in adenosine and noradrenaline receptor density in the rat brain. *Journal of Sleep Research*, *24*(5), 549–558. https://doi.org/10.1111/jsr.12300

Kjellberg A. (1977). Sleep deprivation and some aspects of performance, lapses and other attentional effects. *Waking Sleeping*, *1*, 145–148.

Komada, Y., Asaoka, S., Abe, T., & Inoue, Y. (2013). Short sleep duration, sleep disorders, and traffic accidents. *IATSS Research*, *37*(1), 1–7. https://doi.org/10.1016/j.iatssr.2013.06.001

Kripke, D. F., Garfinkel, L., Wingard, D. L., Klauber, M. R., & Marler, M. R. (2002). Mortality Associated With Sleep Duration and Insomnia. *Archives of General Psychiatry*, *59*(2), 131. https://doi.org/10.1001/archpsyc.59.2.131

Kröger, S., Rutter, B., Stark, R., Windmann, S., Hermann, C., & Abraham, A. (2012). Using a shoe as a plant pot: Neural correlates of passive conceptual expansion. *Brain Research*, *1430*, 52–61. https://doi.org/10.1016/j.brainres.2011.10.031

Kronholm, E., Harma, M., Hublin, C., Aro, A. R., & Partonen, T. (2006). Self-reported sleep duration in Finnish general population. *Journal of Sleep Research*, *15*(3), 276–290. https://doi.org/10.1111/j.1365-2869.2006.00543.x

Kronholm, E., Sallinen, M., Suutama, T., Sulkava, R., Era, P., & Partonen, T. (2009). Self-reported sleep duration and cognitive functioning in the general population. *Journal of Sleep Research*, *18*, 436–446. https://doi.org/10.1111/j.1365-2869.2009.00765.x

Kryger, M. H., Roth, T., & Dement, W. C. (2010). *Principles and Practice of Sleep Medicine E-Book* (5th ed.). Elsevier Health Sciences.

Kuhn, G. (2001). Circadian rhythm, shift work, and emergency medicine. *Annals of Emergency Medicine*, *37*(1), 88–98. https://doi.org/10.1067/mem.2001.111571

Lac, G., & Chamoux, A. (2004). Biological and psychological responses to two rapid shiftwork schedules. *Ergonomics*, *47*(12), 1339–1349. https://doi.org/10.1080/00140130410001724237

Lal, S. K. L., & Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological Psychology*, *55*(3), 173–194. https://doi.org/10.1016/S0301-0511(00)00085-5

Lamond, N., & Dawson, D. (2002). Quantifying the performance impairment associated with fatigue. *Journal of Sleep Research*, *8*(4), 255–262. https://doi.org/10.1046/j.1365-2869.1999.00167.x

Lamond, N., Petrilli, R. M. A., Dawson, D., & Roach, G. D. (2006). Do short international layovers allow sufficient opportunity for pilots to recover? *Chronobiology International*, *23*(6), 1285–1294. https://doi.org/10.1080/07420520601062387

Landrigan, C. P., Rothschild, J. M., Cronin, J. W., Kaushal, R., Burdick, E., Katz, J. T., Lilly, C. M., Stone, P. H., Lockley, S. W., Bates, D. W., & Czeisler, C. A. (2004). Effect of

Reducing Interns' Work Hours on Serious Medical Errors in Intensive Care Units. *New England Journal of Medicine*, *351*(18), 1838–1848. https://doi.org/10.1056/NEJMoa041406

Larouzee, J., & le Coze, J. C. (2020). Good and bad reasons: The Swiss cheese model and its critics. *Safety Science*, *126*, 104660, https://doi.org/10.1016/j.ssci.2020.104660

Leino, T. K., Lohi, J. J., Huttunen, K. H., Lahtinen, T. M., Kilpelainen, A. A., & Muhli, A. A. (2007). Effect of Caffeine on Simulator Flight Performance in Sleep-Deprived Military Pilot Students. *Military Medicine*, *172*(9), 982–987. https://doi.org/Cn-00703901 Pubmed 17937364

Lenné, M. G., Triggs, T. J., & Redman, J. R. (1998). Interactive Effects of Sleep Deprivation, Time of Day, and Driving Experience on a Driving Task. *Sleep*, *21*(1), 38–44. https://doi.org/10.1093/sleep/21.1.38

Leveson, N.G. (2017). *Engineering a Safer World: Systems Thinking applied to Safety*. (2nd ed). MIT Press

Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences*, *1129*, 305–322. https://doi.org/10.1196/annals.1417.002

Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological Bulletin*, *136*(3), 375–389. https://doi.org/10.1037/a0018883

Liu, C. C., Hosking, S. G., & Lenné, M. G. (2009). Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of Safety Research, 40*(4), 239–245. https://doi.org/10.1016/j.jsr.2009.04.005

lo Martire, V., Caruso, D., Palagini, L., Zoccoli, G., & Bastianini, S. (2020). Stress & Sleep: A relationship lasting a lifetime. *Neuroscience and Biobehavioral Reviews*, *117*, 65–77. https://doi.org/10.1016/j.neubiorev.2019.08.024

Lockley, S. W., Barger, L. K., Ayas, N. T., Rothschild, J. M., Czeisler, C. A., & Landrigan, C. P. (2007). Effects of Health Care Provider Work Hours and Sleep Deprivation on Safety and Performance. *The Joint Commission Journal on Quality and Patient Safety, 33*(11), 7–18. https://doi.org/10.1016/S1553-7250(07)33109-7

Lockley, S. W., Cronin, J. W., Evans, E. E., Cade, B. E., Lee, C. J., Landrigan, C. P., Rothschild, J. M., Katz, J. T., Lilly, C. M., Stone, P. H., Aeschbach, D., & Czeisler, C. A. (2004). Effect of Reducing Interns' Weekly Work Hours on Sleep and Attentional Failures. *New England Journal of Medicine, 351*(18), 1829–1837. https://doi.org/10.1056/NEJMoa041404

Loureiro, F., & Garcia-Marques, T. (2015). Morning or Evening person? Which type are you? Self-assessment of chronotype. *Personality and Individual Differences*, *86*, 168–171. https://doi.org/10.1016/j.paid.2015.06.022

Lowden, A., Kecklund, G., Axelsson, J., & Åkerstedt, T. (1998). Change from an 8-hour shift to a 12-hour shift, attitudes, sleep, sleepiness and performance. *Scandinavian Journal of Work, Environment and Health*, *24*(Suppl. 3), 69–75.

Luna, T. D., French, J., & Mitcha, J. L. (1997). A study of USAF air traffic controller shiftwork: sleep, fatigue, activity, and mood analyses. *Aviation, Space, and Environmental Medicine*, *68*(1), 18–23.

Luxhøj, J. T., & Kauffeld, K. (2003). Evaluating the Effect of Technology Insertion into the National Airspace System. *The Rutgers Scholar - An Electronic Bulletin of Undergraduate Research*, *5*.

Luxhøj, J. T., & Maurino, M. (2001). An aviation system risk model (ASRM) case study: Air Ontario 1363. *The Rutgers Scholar*.

Luyster, F. S., Strollo, P. J., Zee, P. C., & Walsh, J. K. (2012). Sleep: A Health Imperative. *Sleep*, *35*(6), 727–734. https://doi.org/10.5665/sleep.1846

Lyman, G. E., & Orlady, H. W. (1981). NASA Fatigue and Associated Performance Decrements in Air Transport Operations. *NASA Technical Reports Server: 19810014163*, 1–32. https://doi.org/10.1111/add.14038

Lynn Caldwell, J., Chandler, J. F., & Hartzler, B. M. (2012). Battling fatigue in aviation: Recent advancements in research and practice. *Journal of Medical Sciences*, *32*(2), 47–56.

Lyznicki, J. M., Doege, T. C., Davis, R. M., & Williams, M. A. (1998). Sleepiness, driving, and motor vehicle crashes. *JAMA*, *279*, 1908–1913.

MacLean, A. W., Davies, D. R. T., & Thiele, K. (2003). The hazards and prevention of driving while sleepy. *Sleep Medicine Reviews*, *7*(6), 507–521. https://doi.org/10.1016/S1087-0792(03)90004-9

Magill, R. A., Waters, W. F., Bray, G. A., Volaufova, J., Smith, S. R., Lieberman, H. R., McNevin, N., & Ryan, D. H. (2003). Effects of tyrosine, phentermine, caffeine D-amphetamine, and placebo on cognitive and motor performance deficits during sleep deprivation. *Nutritional Neuroscience*, *6*(4), 237–246. https://doi.org/10.1080/1028415031000120552

Magnavita, N., & Garbarino, S. (2017). Sleep, health and wellness at work: A scoping review. *International Journal of Environmental Research and Public Health, 14*(11), 1347. https://doi.org/10.3390/ijerph14111347

Mallis, M. M., Maislin, G., Powell, J. W., Staszewski, J. J., Grace, R., & Dinges, D. F. (1998). New drowsiness detection technologies testing their validity to track hypovigilance. *Sleep, 21*(Suppl. 1), 172.

Mallis, M. M., Mejdal, S., Nguyen, T. T., & Dinges, D. F. (2004). Summary of the key features of seven biomathematical models of human fatigue and performance. *Aviation Space and Environmental Medicine*, *75,* 4-14.

Manuele, F. A. (2009). Leading & Lagging Indicators: Do they add value to the practice of safety? *Professional Safety*, 28–33. https://aeasseincludes.assp.org/professionalsafety/pastissues/054/12/F2Manuele_1209.pdf

Maquet, P. (2001). The Role of Sleep in Learning and Memory. *Science, 294*(5544), 1048–1052. https://doi.org/10.1126/science.1062856

Marcus, J. H., & Rosekind, M. R. (2017b). Fatigue in transportation: NTSB investigations and safety recommendations. *Injury Prevention*, *23*(4), 232–238. https://doi.org/10.1136/injuryprev-2015-041791

Martynhak, B. J., Louzada, F. M., Pedrazzoli, M., & Araujo, J. F. (2010). Does the chronotype classification need to be updated? Preliminary findings. *Chronobiology International*, *27*(6), 1329–1334. https://doi.org/10.3109/07420528.2010.490314

Maruff, P., Falleti, M. G., Collie, A., Darby, D., & McStephen, M. (2005). Fatigue-related impairment in the speed, accuracy and variability of psychomotor performance: comparison with blood alcohol levels. *Journal of Sleep Research*, *14*(1), 21–27. https://doi.org/10.1111/j.1365-2869.2004.00438.x

Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology Section A*, *55*(2), 659–686. https://doi.org/10.1080/02724980143000505

McCarley, R. W. (2007). Neurobiology of REM and NREM sleep. *Sleep Medicine, 8*(4), 302–330. https://doi.org/10.1016/j.sleep.2007.03.005

McCauley, P., Kalachev, L. V., Mollicone, D. J., Banks, S., Dinges, D. F., & van Dongen, H. P. A. (2013). Dynamic circadian modulation in a biomathematical model for the effects of sleep and sleep loss on waking neurobehavioral performance. *Sleep*, *36*(12), 1987–1997.

Mcnamara, J. P. H., Wang, J., Holiday, D. B., Warren, J. Y., Paradoa, M., Balkhi, A. M., Fernandez-Baca, J., & McCrae, C. S. (2014). Sleep disturbances associated with cigarette smoking. *Psychology, Health and Medicine*, *19*(4), 410–419. https://doi.org/10.1080/13548506.2013.832782

Mistlberger, R. E., & Skene, D. J. (2004). Social influences on mammalian circadian rhythms: animal and human studies. *Biological Reviews*, *79*(3), 533–556. https://doi.org/10.1017/S1464793103006353

Mitchell, S. J., & Evans, A. D. (2004). Flight Safety and Medical Incapacitation Risk of Airline Pilots. *Aviation Space and Environmental Medicine*, *75*(3 SEC. I), 260–268.

Mitler, M. M., Carskadon, M. A., & Hirshkowitz, M. (2005). Evaluating Sleepiness. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and Practice of Sleep Medicine* (4th ed., pp. 1417–1423). Elsevier. https://doi.org/10.1016/B0-7216-0797-7/X5001-0

Mitler, M. M., Miller, J. C., Lipsitz, J. J., Walsh, J. K., & Wylie, C. D. (1997). The Sleep of Long-Haul Truck Drivers. *New England Journal of Medicine*, *337*(11), 755–762. https://doi.org/10.1056/NEJM199709113371106

Monk, T. H. (1987). Subjective Ratings of Sleepiness—The Underlying Circadian Mechanisms. *Sleep*, *10*(4), 343–353. https://doi.org/10.1093/sleep/10.4.343

Monk, T. H. (1989). A visual analogue scale technique to measure global vigor and affect. *Psychiatry Research*, *27*(1), 89–99. https://doi.org/10.1016/0165-1781(89)90013-9

Moore-Ede, M. (1993). Aviation safety and pilot error. In *Understanding Human Limits in a World That Never Stops: The Twenty four hour society* (pp. 81–96). Wiley.

Mu, Q., Nahas, Z., Johnson, K. A., Yamanaka, K., Mishory, A., Koola, J., Hill, S., Horner, M. D., Bohning, D. E., & George, M. S. (2005). Decreased Cortical Response to Verbal Working Memory Following Sleep Deprivation. *Sleep*, *28*(1), 55–67. https://doi.org/10.1093/sleep/28.1.55

Musico, B. (1921). Is a fatigue test possible? *British Journal of Psychology*, *12*, 31–46.

National Aeronautics and Space Administration, & NASA. (1996). Principles and Guidelines for Duty and Rest Scheduling in Commercial Aviation. *Technical Memorandum*, *110404*, 2–3.

National Research Council. (2011). *The Effects of Commuting on Pilot Fatigue*. National Academies Press. https://doi.org/10.17226/13201

National Transportation Safety Board. (1994). NTSB Aircraft Accident Report: AAR-94/04 Uncontrolled Collision With Terrain. *Aircraft Accident Report*, *AAR-94/04*, 2–78.

National Transportation Safety Board. (2010). *NTSB/AAR-10/01 Accident Report: Loss of Control on Approach Colgan Air Flight 3407*. https://www.ntsb.gov/investigations/reports_aviation.html

National Transportation Safety Board. (2017). *NTSB/AIR-18/01 Incident Report: Taxiway Overflight Air Canada Flight 759*. https://www.ntsb.gov/investigations/AccidentReports/Reports/AIR1801.pdf

National Transportation Safety Board. (2019). *Most Wanted List: Reduce Fatigue-Related Accidents*. Retrieved: https://www.ntsb.gov/safety/mwl/Pages/default.aspx

Neri, D. F., Oyung, R. L., & Dinges, D. (2002). Controlled breaks as a fatigue countermeasure on the flight deck . *Aviation Space and Environmental Medicine* , *73*(7), 654–664.

Neri, D. F., Shappell, S. A., & DeJohn, C. A. (1992). Simulated Sustained Flight Operations and Performance, Part 1: Effects of Fatigue. *Military Psychology*, *4*(3), 137–155. https://doi.org/10.1207/s15327876mp0403_2

Nilson, J. P., Soderstrom, M., Karlsson, A. U., Lekander, M., Åkerstedt, T., Lindroth, N. E., & Axelsson, J. (2005). Less effective executive functioning after one night's sleep deprivation. *Journal of Sleep Research*, *14*(1), 1–6. https://doi.org/10.1111/j.1365-2869.2005.00442.x

Ogilvie, R. D., Mcdonagh, D. M., Stone, S. N., & Wilkinson, R. T. (1988). Eye Movements and the Detection of Sleep Onset. *Psychophysiology*, *25*(1), 81–91. https://doi.org/10.1111/j.1469-8986.1988.tb00963.x

Ogilvie, R. D., Wilkinson, R. T., & Allison, S. (1989). The detection of sleep onset: behavioral, physiological, and subjective convergence. *Sleep*, *12*(5), 458–474.

O'Hagan, A. D., Issartel, J., Fletcher, R., & Warrington, G. (2016). Duty hours and incidents in flight among commercial airline pilots. *International Journal of Occupational Safety and Ergonomics*, *22*(2), 165–172. https://doi.org/10.1080/10803548.2016.1146441

O'Hagan, A. D., Issartel, J., Wall, A., Dunne, F., Boylan, P., Groeneweg, J., Herring, M., Campbell, M., & Warrington, G. (2020). "Flying on empty" – effects of sleep deprivation on pilot performance. *Biological Rhythm Research*, *51*(7), 1133–1154. https://doi.org/10.1080/09291016.2019.1581481

Onninen, J., Hakola, T., Puttonen, S., Tolvanen, A., Virkkala, J., & Sallinen, M. (2020). Sleep and sleepiness in shift-working tram drivers. *Applied Ergonomics*, *88*, 103153.

Otmani, S., Pebayle, T., Roge, J., & Muzet, A. (2005). Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers. *Physiology & Behavior*, *84*(5), 715–724. https://doi.org/10.1016/j.physbeh.2005.02.021

Pack, A. I., Pack, A. M., Rodgman, E., Cucchiara, A., Dinges, D. F., & Schwab, C. W. (1995). Characteristics of crashes attributed to the driver having fallen asleep. *Accident Analysis & Prevention*, *27*(6), 769–775. https://doi.org/10.1016/0001-4575(95)00034-8

Paley, M. J., & Tepas, D. I. (1994). Fatigue and the Shiftworker: Firefighters Working on a Rotating Shift Schedule. *Human Factors*, *36*(2), 269–284.

Pascoe, P. A., Johnson, M. K., Robertson, K. A., & Spencer, M. B. (1995). *Sleep in rest facilities on-board aircraft: field studies*. DRA Report No. DRA/CHS/A&N/CR/95/002

Patriarca, R., Di Gravio, G., Woltjer, R., Costantino, F., Praetorius, G., Ferreira, P., & Hollnagel, E. (2020). Framing the FRAM: A literature review on the functional resonance analysis method. *Safety Science*, *129*, 104827.

Patrick, G. T. W., & Gilbert, J. A. (1896). Studies from the psychological laboratory of the University of Iowa: On the effects of loss of sleep. *Psychological Review*, *3*(5), 469–483. https://doi.org/10.1037/h0075739

Patton, D. F., & Mistlberger, R. E. (2013). Circadian adaptations to meal timing: Neuroendocrine mechanisms. In *Frontiers in Neuroscience* (Issue 7 OCT). https://doi.org/10.3389/fnins.2013.00185

Petrie, K. J., Powell, D., & Broadbent, E. (2004). Fatigue self-management strategies and reported fatigue in international pilots. *Ergonomics*, *47*, 461–468. https://doi.org/10.1080/0014013031000085653

Petrilli, R. M., Roach, G. D., Dawson, D., & Lamond, N. (2006). The sleep, subjective fatigue, and sustained attention of commercial airline pilots during an international pattern. *Chronobiology International*, *23*(6), 1357–1362. https://doi.org/10.1080/07420520601085925

Philip, P., Sagaspe, P., Moore, N., Taillard, J., Charles, A., Guilleminault, C., & Bioulac, B. (2005). Fatigue, sleep restriction and driving performance. *Accident Analysis & Prevention*, *37*(3), 473–478. https://doi.org/10.1016/j.aap.2004.07.007

Philip, P., Sagaspe, P., Prague, M., Tassi, P., Capelli, A., Bioulac, B., Commenges, D., & Taillard, J. (2012). Acute Versus Chronic Partial Sleep Deprivation in Middle-Aged People: Differential Effect on Performance and Sleepiness. *Sleep*, *35*(7), 997-1002. https://doi.org/10.5665/sleep.1968

Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Åkerstedt, T., Charles, A., & Bioulac, B. (2005). Fatigue, Sleepiness, and Performance in Simulated Versus Real Driving Conditions. *Sleep*, *28*(12), 1511–1516.

Phillips, R. O. (2015). A review of definitions of fatigue - And a step towards a whole definition. *Transportation Research Part F: Traffic Psychology and Behaviour*, *29*, 48–56. https://doi.org/10.1016/j.trf.2015.01.003

Phillips, R. O., Kecklund, G., Anund, A., & Sallinen, M. (2017). Fatigue in transport: a review of exposure, risks, checks and controls. *Transport Reviews*, *37*(6), 742–766. https://doi.org/10.1080/01441647.2017.1349844

Pilcher, J. J., Band, D., Odle-Dusseau, H. N., & Muth, E. R. (2007). Human performance under sustained operations and acute sleep deprivation conditions: toward a model of controlled attention. *Aviation, Space, and Environmental Medicine*, *78*(5), 15–24.

Pilcher, J. J., Geldhauser, H. A., Beeco, J. A., & Lindquist, T. A. (2013). Controlled attention and sleep deprivation: adding a self-regulation approach? *International Journal of Psychological Studies*, *5*(3), 71–79.

Pilcher, J. J., & Huffcutt, A. (1996). Effects of Sleep Deprivation on Performance: A Meta-Analysis. *Sleep*, *19*(4), 318–326. https://academic.oup.com/sleep/article/19/4/318/2749842

Pilcher, J. J., Lambert, B. J., & Huffcutt, A. I. (2000). Differential Effects of Permanent and Rotating Shifts on Self-Report Sleep Length: A Meta-Analytic Review. *Sleep*, *23*(2), 1–8.

Pires, M. L. N., Teixeira, C. W., Esteves, A. M., Bittencourt, L. R. A., Silva, R. S., Santos, R. F., Tufik, S., & Mello, M. T. (2009). Sleep, ageing and night work. *Brazilian Journal of Medical and Biological Research*, *42*(9), 839–843.

Porkka-Heiskanen, T., Alanko, L., Kalinchuk, A., & Stenberg, D. (2002). Adenosine and sleep. *Sleep Medicine Reviews*, *6*(4), 321–332. https://doi.org/10.1053/smrv.2001.0201

Porkka-Heiskanen, T., Strecker, R. E., & McCarley, R. W. (2000). Brain site-specificity of extracellular adenosine concentration changes during sleep deprivation and spontaneous sleep: an in vivo microdialysis study. *Neuroscience*, *99*(3), 507–517. https://doi.org/10.1016/S0306-4522(00)00220-7

Pouliquen, Y., Ferrante, O., Jouniaux, G., Cabon, P., Rome, F., Wolff, M., & Mollard, R. (2005). A Human Factors Approach for the Analysis and the Encoding of Aviation Accidents and Incidents: A Validation Study. *2005 International Symposium on Aviation Psychology*, 578–583.

Powell, D. M. C., Spencer, M. B., Holland, D., Broadbent, E., & Petrie, K. J. (2007). Pilot fatigue in short-haul operations: Effects of number of sectors, duty length, and time of day. *Aviation Space and Environmental Medicine*, *78*(7), 698–701.

Powell, D. M. C., Spencer, M. B., & Petrie, K. J. (2011). Automated collection of fatigue ratings at the top of descent: A practical commercial airline tool. *Aviation Space and Environmental Medicine*, *82*(11), 1037–1041. https://doi.org/10.3357/ASEM.3115.2011

Powell, D., Spencer, M. B., Holland, D., & Petrie, K. J. (2008). Fatigue in two-pilot operations: Implications for flight and duty time limitations. *Aviation Space and Environmental Medicine*, *79*, 1047–1050. https://doi.org/10.3357/ASEM.2362.2008

Prabhu, M. S. (2021). Design of Microsleep Alerting System of Pilot to Reduce Air Accidents. *International Journal of Engineering and Management Research*, *11*(1), 42–45. https://doi.org/10.31033/ijemr.11.1.6

Previc, F. H., Lopez, N., Ercoline, W. R., Daluz, C. M., Workman, A. J., Evans, R. H., & Dillon, N. A. (2009). The Effects of Sleep Deprivation on Flight Performance, Instrument Scanning, and Physiological Arousal in Pilots. *The International Journal of Aviation Psychology*, *19*(4), 326–346. https://doi.org/10.1080/10508410903187562

Provan, D. J., Woods, D. D., Dekker, S. W., & Rae, A. J. (2020). Safety II professionals: How resilience engineering can transform safety practice. *Reliability Engineering & System Safety*, *195*, 106740.

Quigley, N., Green, J. F., Morgan, D., Idzikowski, C., & King, D. J. (2000). The effect of sleep deprivation on memory and psychomotor function in healthy volunteers. *Human Psychopharmacology: Clinical and Experimental*, *15*(3), 171–177. https://doi.org/10.1002/(SICI)1099-1077(200004)15:3<171::AID-HUP155>3.0.CO;2-D

Rail Accident Investigation Branch (RAIB). (2020). *Report 18/2017: Overturning of a tram at Sandilands junction, Croydon*. Retrieved 05/04/2022: https://assets.publishing.service.gov.uk/media/5de79643e5274a06dee23a10/R182017_201022_Sandilands_v2.2.pdf

Railways and Transport Act, UK Legislation (2003). Retrieved 05/07/2019: https://www.legislation.gov.uk/ukpga/2003/20/contents

Rajaratnam, S. M. W., Middleton, B., Stone, B. M., Arendt, J., & Dijk, D. J. (2004). Melatonin advances the circadian timing of EEG sleep and directly facilitates sleep without altering its duration in extended sleep opportunities in humans. *Journal of Physiology*, *561*(1), 339–351. https://doi.org/10.1113/jphysiol.2004.073742

Ramakrishnan, S., Wesensten, N. J., Kamimori, G. H., Moon, J. E., Balkin, T. J., & Reifman, J. (2016). A unified model of performance for predicting the effects of sleep and caffeine. *Sleep, 39*(10), 1827–1841. https://doi.org/10.5665/sleep.6164

Ratcliff, R., & van Dongen, H. P. A. (2009). Sleep deprivation affects multiple distinct cognitive processes. *Psychonomic Bulletin and Review*, *16*(4), 742–751. https://doi.org/10.3758/PBR.16.4.742

Reader, T.W., Parand, A. & Kirwan, B. (2018). *Mapping safety culture onto processes and practices: the Safety Culture Stake Approach.* European Commission Future Sky Safety Report: D5.13, 3-39.

Reason, J. (1990). *Human error*. Cambridge University Press.

Reason, J. (1995). A systems approach to organisational error. *Ergonomics*, *38*(8), 1708–1721.

Reason, J. (2000). Human Error: models and management. *BMJ, 320*, 768–770.

Reason, J., Hollnagel, E., & Paries, J. (2006). Revisiting the ‘Swiss cheese’ model of accidents. *Eurocontrol Experimental Centre*, *EEC Note*(13/06).

Reinke, L., Özbay, Y., Dieperink, W., & Tulleken, J. E. (2015). The effect of chronotype on sleepiness, fatigue, and psychomotor vigilance of ICU nurses during the night shift. *Intensive Care Medicine*, *41*(4), 657–666. https://doi.org/10.1007/s00134-015-3667-7

Reis, C., Mestre, C., & Canhão, H. (2013). Prevalence of fatigue in a group of airline pilots. *Aviation Space and Environmental Medicine, 84*, 828–833. https://doi.org/10.3357/ASEM.3548.2013

Reis, C., Mestre, C., Canhão, H., Gradwell, D., & Paiva, T. (2016). Sleep complaints and fatigue of airline pilots. *Sleep Science*, *9*(2), 73–77. https://doi.org/10.1016/j.slsci.2016.05.003

Reyner, L., & Horne, J. (1998). Falling asleep whilst driving: are drivers aware of prior sleepiness? *International Journal of Legal Medicine*, *111*, 120–123.

Ribeiro, J. A., & Sebastio, A. M. (2010). Caffeine and adenosine. *Journal of Alzheimer’s Disease*, *20*(SUPPL.1). https://doi.org/10.3233/JAD-2010-1379

Riethmeister, V., Matthews, R. W., Dawson, D., de Boer, M. R., Brouwer, S., & Bültmann, U. (2019). Time-of-day and days-on-shift predict increased fatigue over two-week offshore day-shifts. *Applied Ergonomics*, *78*, 157–163. https://doi.org/10.1016/j.apergo.2019.02.010

Roach, G. D., Darwent, D., & Dawson, D. (2010). How well do pilots sleep during long-haul flights? *Ergonomics*, *53*(9), 1072–1075. https://doi.org/10.1080/00140139.2010.506246

Roach, G. D., Sargent, C., Darwent, D., & Dawson, D. (2012). Duty periods with early start times restrict the amount of sleep obtained by short-haul airline pilots. *Accident Analysis and Prevention*, *45*(SUPPL.), 22–26. https://doi.org/10.1016/j.aap.2011.09.020

Rodrigues, T. E., Fischer, F. M., Bastos, E. M., Baia, L., Bocces, R., Gonçalves, F. P., Licati, P. R., Menquini, A., Spyer, P., Stefenon, E., & Helene, A. F. (2020). Seasonal variation in fatigue indicators in Brazilian civil aviation crew rosters. *Revista Brasileira de Medicina Do Trabalho*, *18*(1), 2–10. https://doi.org/10.5327/Z1679443520200467

Roehrs, T., & Roth, T. (2001). Sleep, sleepiness, and alcohol use. *Alcohol Research & Health*, *25*(2), 101–109.

Roelen, A., & van Dijk, H. (2019). *CAP1756. Task 3: Analysis of accidents and incidents involving flight crew fatigue*. Retrieved: http://publicapps.caa.co.uk/docs/33/Pilot%20Fatigue%20Measurement%20Research_Final%20Report_v0.9.1clean.pdf

Roenneberg, T. (2012). What is chronotype? *Sleep and Biological Rhythms*, *10*(2), 75–76. https://doi.org/10.1111/j.1479-8425.2012.00541.x

Roenneberg, T., Kantermann, T., Juda, M., Vetter, C., & Allebrandt, K. v. (2013). Light and the human circadian clock. *Handbook of Experimental Pharmacology*, *217*, 311–331.

Roenneberg, T., Kuehnle, T., Juda, M., Kantermann, T., Allebrandt, K., Gordijn, M., & Merrow, M. (2007). Epidemiology of the human circadian clock. *Sleep Medicine Reviews*, *11*(6), 429–438. https://doi.org/10.1016/j.smrv.2007.07.005

Rosa, D. E., Marot, L. P., de Mello, M. T., Marqueze, E. C., Narciso, F. V., de Araújo, L. B., & Crispim, C. A. (2021). Association between chronotype and psychomotor performance of rotating shift workers. *Scientific Reports*, *11*(1), 6919 https://doi.org/10.1038/s41598-021-86299-8

Rosekind, M. R., Co, E. L., Gregory, K. B., & Miller, D. L. (2000). *Crew Factors in Flight Operations XIII: A Survey of Fatigue Factors in Corporate/Executive Aviation Operations*. Retrieved: http://www.sti.nasa.gov

Rosekind, M. R., Curtis Graeber, R., Dinges, D. F., Rountree, M. S., Spinweber, C. L., & Gillen, K. A. (1994). *Crew Factors in Flight Operations IX: Effects of Planned Cockpit Rest on Crew Performance and Alertness in Long-Haul Operations*.

Rosekind, M. R., Miller, D. L., Gregory, K. B., Software, S., Smith, R. M., Weldon, K. J., Co, E. L., & Mcnally, K. L. (1994). Fatigue in Operational Settings: Examples from the Aviation Environment. *Human Factors*, *36*(2), 327-338.

Rosekind, M., Smith, R., Miller, D., Co, E., Gregory, K., Webbon, L., Gander, P., & Lebacqz, V. (1995). Alertness management: strategic naps in operational settings. *Journal of Sleep Research*, *4*, 62–66. https://doi.org/10.1111/j.1365-2869.1995.tb00229.x

Rosenthal, L., Roehrs, T. A., & Roth, T. (1993). The sleep-wake activity inventory: A self-report measure of daytime sleepiness. *Biological Psychiatry*, *34*(11), 810–820. https://doi.org/10.1016/0006-3223(93)90070-T

Rudari, L., Johnson, M., Geske, R., & Sperlak, L. (2016). Pilot Perceptions on Impact of Crew Rest Regulations on Safety and Fatigue. *International Journal of Aviation, Aeronautics, and Aerospace*. https://doi.org/10.15394/ijaaa.2016.1096

Russo, M. B., Escolas, S., & Thorne, D. (2004). Occurrence and patterns of visual neglect in US Air Force pilots in a simulated overnight flight. *Aviation Space and Environmental Medicine*, *75*(4).

Sagaspe, P., Taillard, J., Åkersted, T., Bayon, V., Espié, S., Chaumet, G., Bioulac, B., & Philip, P. (2008). Extended driving impairs nocturnal driving performances. *PLoS ONE*, *3*(10), 1–6. https://doi.org/10.1371/journal.pone.0003493

Sallinen, M., & Hublin, C. (2015). Fatigue-Inducing Factors in Transportation Operators. *Reviews of Human Factors and Ergonomics*, *10*(1), 138–173. https://doi.org/10.1177/1557234X15574828

Sallinen, M., & Kecklund, G. (2010). Shift work, sleep, and sleepiness - Differences between shift schedules and systems. *Scandinavian Journal of Work, Environment and Health*, *36*(2), 121–133). Nordic Association of Occupational Safety and Health. https://doi.org/10.5271/sjweh.2900

Sallinen, M., Sihvola, M., Puttonen, S., Ketola, K., Tuori, A., Härmä, M., Kecklund, G., & Åkerstedt, T. (2017). Sleep, alertness and alertness management among commercial airline pilots on short-haul and long-haul flights. *Accident Analysis and Prevention*, *98*, 320–329. https://doi.org/10.1016/j.aap.2016.10.029

Sallinen, M., van Dijk, H., Aeschbach, D., Maij, A., & Åkerstedt, T. (2020). A large-scale European Union study of aircrew fatigue during long night and disruptive duties. *Aerospace Medicine and Human Performance*, *91*(8), 628–635. https://doi.org/10.3357/AMHP.5561.2020

Sallinen, M., van Dijk, H., Aeschbach, D., Maij, A., & Åkerstedtm Torbjorn. (2020). A Large-Scale European Union Study of Aircrew Fatigue During Long Night and Disruptive Duties. *Aerospace Medicine and Human Performance*, *91*(8), 628–635.

Samel, A. , Vejvoda, M. , & Maass, H. , 2004. (2004). Sleep deficit and stress hormones in helicopter pilots on 7-day duty for emergency medical services. *Aviation Space and Environmental Medicine*, *75*(11), 935–940.

Samel A, & Wegmann H. (1988). Sleep and circadian rhythms of an airline pilot operating on the transmeridien polar route: a case study. *Aviation, Space and Environmental Medicine, 59*, 443–447.

Samel, A., Wegmann, H. M., & Vejvoda, M. (1997). Aircrew fatigue in long-haul operations. *Accident Analysis and Prevention*, *29*(4 SPEC. ISS.), 439–452.

Sandberg, D., Anund, A., Fors, C., Kecklund, G., Karlsson, J. G., Wahde, M., & Åkerstedt, T. (2011). The Characteristics of Sleepiness During Real Driving at Night—A Study of Driving Performance, Physiology and Subjective Experience. *Sleep*, *34*(10), 1317–1325. https://doi.org/10.5665/sleep.1270

Santamaria, J., & Chiappa, K. H. (1987). The EEG of drowsiness in normal adults. *Journal of Clinical Neurophysiology*, *4*(4), 327–382.

Scheer, F. A. J. L., Wright, K. P., Kronauer, R. E., & Czeisler, C. A. (2007). Plasticity of the Intrinsic Period of the Human Circadian Timing System. *PLoS ONE*, *2*(8), e721. https://doi.org/10.1371/journal.pone.0000721

Schoutens, A. M. C., Brown, L., & Whitehurst, G. (2014). The effects of light exposure on flight crew alertness levels to enhance fatigue risk management prediction models. In S. Losinski, R. Diehl, T. Booker, & T. Davis (Eds.), *Probabilistic Safety Assessment and Management PSAM 12*.

Senaratna, C. v., Perret, J. L., Lodge, C. J., Lowe, A. J., Campbell, B. E., Matheson, M. C., Hamilton, G. S., & Dharmage, S. C. (2017). Prevalence of obstructive sleep apnea in the general population: A systematic review. In *Sleep Medicine Reviews* (Vol. 34, pp. 70–81). W.B. Saunders Ltd. https://doi.org/10.1016/j.smrv.2016.07.002

Shahid, A., Shen, J., & Shapiro, C. M. (2010). Measurements of sleepiness and fatigue. *Journal of Psychosomatic Research, 69*(1), 81–89. https://doi.org/10.1016/j.jpsychores.2010.04.001

Shekari Soleimanloo, S., White, M. J., Garcia-Hansen, V., & Smith, S. S. (2017). The effects of sleep loss on young drivers' performance: A systematic review. *PLOS ONE, 12*(8), e0184002. https://doi.org/10.1371/journal.pone.0184002

Shen, J., Barbera, J., & Shapiro, C. M. (2006). Distinguishing sleepiness and fatigue: Focus on definition and measurement. *Sleep Medicine Reviews, 10*(1), 63–76. https://doi.org/10.1016/j.smrv.2005.05.004

Short, M. A., & Louca, M. (2015). Sleep deprivation leads to mood deficits in healthy adolescents. *Sleep Medicine*, *16*(8), 987–993. https://doi.org/10.1016/j.sleep.2015.03.007

Signal, L., Ratieta, D., & Gander, P. (2006). *ATSB Research and Analysis Report: Fatigue Management in the New Zealand Aviation Industry*.

Signal, T. L., Gander, P. H., van den Berg, M. J., & Graeber, R. C. (2013a). In-Flight Sleep of Flight Crew During a 7-hour Rest Break: Implications for Research and Flight Safety. *Sleep*, *36*(1), 109–115. https://doi.org/10.5665/sleep.2312

Smith, A., & Maben, A. (1993). Effects of sleep deprivation, lunch, and personality on performance, mood, and cardiovascular function. *Physiology & Behavior*, *54*(5), 967–972. https://doi.org/10.1016/0031-9384(93)90310-C

Smith, M. E., Mcevoy, L. K., & Gevins, A. (2002). The Impact of Moderate Sleep Loss on Neurophysiologic Signals during Working-Memory Task Performance. *Sleep*, *25*(7), 784–794.

Smolensky, M. H., di Milia, L., Ohayon, M. M., & Philip, P. (2011). Sleep disorders, medical conditions, and road accident risk. *Accident Analysis and Prevention*, *43*(2), 533–548. https://doi.org/10.1016/j.aap.2009.12.004

Sneddon, A., Mearns, K., & Flin, R. (2013). Stress, fatigue, situation awareness and safety in offshore drilling crews. *Safety Science*, *56*, 80–88. https://doi.org/10.1016/j.ssci.2012.05.027

Söderström, M., Ekstedt, M., & Åkerstedt, T. (2006). Weekday and weekend patterns of diurnal cortisol, activation and fatigue among people scoring high for burnout. *Scandinavian Journal of Work, Environment & Health*, *2*(2), 35-40.

Spencer, M. B., & Montgomery, J. M. (1997). *Sleep patterns of aircrew on charter / air haulage routes*. DERA Report: Safety Regulation Group of the CAA.

Spencer, M. B., & Robertson, K. A. (2000). *Report No DERA/CHS/PPD/CR0000: A diary study of aircrew fatigue in short-haul multi-sector operations*. Retrieved 21/09/2020: from https://publicapps.caa.co.uk/docs/33/CAAPaper2005_04.pdf

Spencer, M. B., & Robertson, K. A. (2002). *Aircrew alertness during short-haul operations, including the impact of early starts*. Retrieved 18/05/2020: https://www.eurocockpit.be/sites/default/files/Spencer-Robertson-2002.pdf

Sprajcer, M., Thomas, M. J. W., Sargent, C., Crowther, M. E., Boivin, D. B., Wong, I. S., Smiley, A., & Dawson, D. (2022). How effective are Fatigue Risk Management Systems

(FRMS)? A review. *Accident Analysis & Prevention*, *165*, 106398. https://doi.org/10.1016/j.aap.2021.106398

Steptoe, A., & Bostock, S. (2011). *A survey of fatigue and well-being among commercial airline pilots.* UCL Psychobiology Group, London

Steiner, S., Fakleš, D., & Gradišar, T. (2012). Problems of crew fatigue management in airline operations. *Transport Engineering*, *617*, 626-623.

Stern, J. A., Walrath, L. C., & Goldstein, R. (1984). The Endogenous Eyeblink. *Psychophysiology*, *21*(1), 22–33. https://doi.org/10.1111/j.1469-8986.1984.tb02312.x

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1272–1278. https://doi.org/10.1038/nature04286

Stolzer, A.J. & Goglia, J.J. (2015). *Safety Management Systems in Aviation.* (2nd ed.). New York: Routledge

Swiss Aircraft Accident Investigation Bureau. (2001). *SAIB Final Report No. 1793*. 1–58.

Tarokh, L., Rusterholz, T., Achermann, P., & van Dongen, H. P. A. (2015). The spectrum of the non-rapid eye movement sleep electroencephalogram following total sleep deprivation is trait-like. *Journal of Sleep Research*, *24*(4), 360–363. https://doi.org/10.1111/jsr.12279

Tefft, B. C. (2018). Acute sleep deprivation and culpable motor vehicle crash involvement. *Sleep*, *41*(10). https://doi.org/10.1093/sleep/zsy144

Tembo, A. C., & Parker, V. (2009). Factors that impact on sleep in intensive care patients. *Intensive and Critical Care Nursing*, *25, 6*, 314–322. https://doi.org/10.1016/j.iccn.2009.07.002

Tempesta, D., Couyoumdjian, A., Curcio, G., Moroni, F., Marzano, C., de Gennaro, L., & Ferrara, M. (2010). Lack of sleep affects the evaluation of emotional stimuli. *Brain Research Bulletin*, *82*(1–2), 104–108. https://doi.org/10.1016/j.brainresbull.2010.01.014

Tepas, D. I., Armstrong, D. R., Carlson, M. L., Duchon, J. C., Gersten, A., & Lezotte, D. v. (1985). Changing industry to continuous operations: Different strokes for different plants. *Behavior Research Methods, Instruments, & Computers*, *17*(6), 670-676.

Terman, M., & Terman, J. S. (2005). Light Therapy. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and Practice of Sleep Medicine* (4th ed., pp. 1424–1442). Elsevier. www.cet.org.

Thomas, L. C., Gast, C., Grube, R., & Craig, K. (2015). Fatigue Detection in Commercial Flight Operations: Results Using Physiological Measures. *Procedia Manufacturing*, *3*, 2357–2364. https://doi.org/10.1016/j.promfg.2015.07.383

Thomas, M. J. W., Petrilli, R. M., Lamond, N. A., Dawson, D., & Roach, G. D. (2006). Australian Long Haul Fatigue Study. *Enhancing Safety Worldwide: Proceedings of the 59th Annual International Air Safety Seminar*.

Tilley, A. J., Wilkinson, R. T., Warren, P. S. G., Watson, B., & Drud, M. (1982). The Sleep and Performance of Shift Workers. *Human Factors*, *24*(6), 629-641. https://doi.org/10.1177/001872088202400601

Tilley, B., & Brown, S. (1992). Chapter 9. Sleep Deprivation. In A. P. Smith & D. M. Jones (Eds.), *Handbook of Human Performance* (pp. 237–260). Academic Press: Harcourt Brace Janovich.

Torsvall, L., & Åkerstedt, T. (1987). Sleepiness on the job: continuously measured EEG changes in train drivers. *Electroencephalography and Clinical Neurophysiology*, *66*(6), 502–511. https://doi.org/10.1016/0013-4694(87)90096-4

Torsvall, L., Åkerstedt, T., Gillander, K., & Knutsson, A. (1989). Sleep on the Night Shift: 24-Hour EEG Monitoring of Spontaneous Sleep/Wake Behavior. *Psychophysiology*, *26*(3), 352–358. https://doi.org/10.1111/j.1469-8986.1989.tb01934.x

Trutschel, U., Sirois, B., Sommer, D., Golz, M., & Edwards, D. (2011). PERCLOS: An alertness measure of the past. *Driving Assessment Conference (Vol. 6, No. 2011)*.

TUC. (2018). *Trade Union Communication Awards*. Retrieved 02/11/2019: https://www.tuc.org.uk/sites/default/files/Comms_Awards_2018_Booklet_AW_Digital_Optimised.pdf

Tunstall-Pedoe, H. (1988). The Concept of Risk. *European Heart Journal*, *9*, 13–15. https://doi.org/10.1111/j.2041-6962.1978.tb01330.x

Underwood, P., & Waterson, P. (2014). Systems thinking, the Swiss Cheese Model and accident analysis: a comparative systemic analysis of the Grayrigg train derailment using the ATSB, AcciMap and STAMP models. *Accident Analysis & Prevention*, *68*, 75–94.

van de Ven, H. A., van der Klink, J. J. L., Vetter, C., Roenneberg, T., Gordijn, M., Koolhaas, W., de Looze, M. P., Brouwer, S., & Bültmann, U. (2016). Sleep and need for recovery in shift workers: do chronotype and age matter? *Ergonomics*, *59*(2), 310–324. https://doi.org/10.1080/00140139.2015.1058426

van Dongen, H. P. (2006). Shift work and inter-individual differences in sleep and sleepiness. *Chronobiology International*, *23*(6), 1139–1147. https://doi.org/10.1080/07420520601100971 van Dongen, H. P. A. (2004). Comparison of mathematical model predictions to experimental data of fatigue and performance. *Aviation Space and Environmental Medicine*, *75*, 15-36.

van Dongen, H. P. A., & Belenky, G. (2009). Individual Differences in Vulnerability to Sleep Loss in the Work Environment. *Industrial Health*, *47*, 518–526.

van Dongen, H. P. A., Bender, A. M., & Dinges, D. F. (2012). Systematic individual differences in sleep homeostatic and circadian rhythm contributions to neurobehavioral impairment during sleep deprivation. *Accident Analysis and Prevention*, *45*(SUPPL.), 11–16. https://doi.org/10.1016/j.aap.2011.09.018

van Dongen, H. P. A., Caldwell, J. A., & Caldwell, J. L. (2011). Individual differences in cognitive vulnerability to fatigue in the laboratory and in the workplace. *Progress in Brain Research*, *190*, 145–153. https://doi.org/10.1016/B978-0-444-53817-8.00009-8

van Dongen, H. P. A., & Dinges, D. (2005). Circadian rhythms in Sleepiness, Alertness, and Performance. In Kryger M, Roth T, & Dement W (Eds.), *Principles and Practice of Sleep Medicine* (4th ed., pp. 435–443). Elsevier.

van Dongen, H. P. A., Maislin, G., & Dinges, D. F. (2004). Dealing with inter-individual differences in the temporal dynamics of fatigue and performance: Importance and techniques. *Aviation Space and Environmental Medicine*, *75*.

van Dongen, H. P. A., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness. *Sleep (Rochester)*, *26*, 117–126. https://doi.org/10.1093/sleep/26.2.117

van Dongen, H. P. A., Rogers, N. L., & Dinges, D. F. (2003). Sleep debt: Theoretical and Empirical Issues (Review). *Sleep and Biological Rhythms*, *1*(1), 5–13.

van Reeth, O., Weibel, L., Spiegel, K., Leproult, R., Dugovic, C., & Maccari, S. (2000). Interactions between stress and sleep: From basic research to clinical situations. *Sleep Medicine Reviews*, *4*(2), 201–219. https://doi.org/10.1053/smrv.1999.0097

Vartanian, O., Bouak, F., Caldwell, J. L., Cheung, B., Cupchik, G., Jobidon, M.-E., Lam, Q., Nakashima, A., Paul, M., Peng, H., Silvia, P. J., & Smith, I. (2014). The effects of a single night of sleep deprivation on fluency and prefrontal cortex function during divergent thinking. *Frontiers in Human Neuroscience*, *8*, 1-12. https://doi.org/10.3389/fnhum.2014.00214

Venus, M. (2021). Short and Long Haul Pilots' Rosters, Stress, Sleep Problems, Fatigue, Mental Health, and Well-Being. *Aerospace Medicine and Human Performance*, *92*(10), 786–797. https://doi.org/10.3357/AMHP.5812.2021

Venus, M., & Holtforth, M. grosse. (2021). Short and Long Haul Pilots Rosters, Stress, Sleep Problems, Fatigue, Mental Health, and Well-Being. *Aerospace Medicine and Human Performance*, *92*(10), 786–797. https://doi.org/10.3357/AMHP.5812.2021

Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annual Review of Psychology*, *57*, 139–166. https://doi.org/10.1146/annurev.psych.56.091103.070307

Waters, F., & Bucks, R. S. (2011). Neuropsychological effects of sleep loss: Implication for neuropsychologists. *Journal of the International Neuropsychological Society*, *17*(4), 571–586. https://doi.org/10.1017/S1355617711000610

Watt, T. (2000). Fatigue in the Danish general population. Influence of sociodemographic factors and disease. *Journal of Epidemiology & Community Health*, *54*(11), 827–833. https://doi.org/10.1136/jech.54.11.827

Wewers, M. E., & Lowe, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health*, *13*(4), 227–236. https://doi.org/10.1002/nur.4770130405

Wilkinson, R. T. (1961). Interaction of lack of sleep with knowledge of results, repeated testing, and individual difference. *Journal of Psychology*, *62*, 263–271.

Williams, H. L., Lubin, A., & Goodnow, J. (1959). Impaired performance with acute sleep loss. *Psychological Monographs*, *73*(14), 1-26.

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*(2), 498–515. https://doi.org/10.1016/j.aap.2009.11.011

Williamson, A. M., & Feyer, A. M. (2000). Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occupational and Environmental Medicine*, *57*, 649–655.

Williamson, A. M., & Sanderson, J. W. (1986). Changing the speed of shift rotation: a field study. *Ergonomics*, *29*(9), 1085–1096. https://doi.org/10.1080/00140138608967225

Wilson, G. F., Caldwell, J. A., & Russell, C. A. (2007). Performance and Psychophysiological Measures of Fatigue Effects on Aviation Related Tasks of Varying Difficulty. *The International Journal of Aviation Psychology*, *17*(2), 219–247. https://doi.org/10.1080/10508410701328839

Wilson, G. F., Russell, C., & Caldwell, J. (2006). Performance and Psychophysiological Measures of Fatigue Effects on Aviation Related Tasks of Varying Difficulty. *International Journal of Aviation Psychology*, *17*(2), 51. DOI:10.1080/10508410701328839

Wilson, M., Permito, R. , English, A., Albritton, S., Coogle, C., & van Dongen, H. P. (2019). Performance and sleepiness in nurses working 12-h day shifts or night shifts in a community hospital. *Accident Analysis and Prevention*, *126*, 43–46.

Wimmer, F., Hoffmann, R. F., Bonato, R. A., & Moffitt, A. R. (1992). The effects of sleep deprivation on divergent thinking and attention processes. *Journal of Sleep Research*, *1*(4), 223–230. https://doi.org/10.1111/j.1365-2869.1992.tb00043.x

Wingelaar-Jagt, Y. Q., Wingelaar, T. T., Riedel, W. J., & Ramaekers, J. G. (2021). Fatigue in Aviation: Safety Risks, Preventive Strategies and Pharmacological Interventions. In *Frontiers in Physiology* (Vol. 12). Frontiers Media S.A. https://doi.org/10.3389/fphys.2021.712628

Wong, K. K. H., Marshall, N. S., Grunstein, R. R., Dodd, M. J., & Rogers, N. L. (2008). Comparing the neurocognitive effects of 40 h sustained wakefulness in patients with untreated OSA and healthy controls. *Journal of Sleep Research*, *17*(3), 322–330. https://doi.org/10.1111/j.1365-2869.2008.00665.x

Wright, K. P., Bogan, R. K., & Wyatt, J. K. (2013). Shift work and the assessment and management of shift work disorder (SWD). *Sleep Medicine Reviews*, *17*(1), 41–54. https://doi.org/10.1016/j.smrv.2012.02.002

Wright, N., & McGown, A. (2001). Vigilance on the civil flight deck: Incidence of sleepiness and sleep during long-haul flights and associated changes in physiological parameters. *Ergonomics*, *44*(1), 82–106. https://doi.org/10.1080/00140130150203893

Wright, N., & McGown, A. (2004). Involuntary Sleep during Civil Air Operations: Wrist Activity and the Prevention of Sleep. *Aviation Space and Environmental Medicine*, *75*(1), 37–45.

Wu, J. C., Gillin, J. C., Buchsbaum, M. S., Chen, P., Keator, D. B., Khosla Wu, N., Darnall, L. A., Fallon, J. H., & Bunney, W. E. (2006). Frontal lobe metabolic decreases with sleep deprivation not totally reversed by recovery sleep. *Neuropsychopharmacology*, *31*(12), 2783–2792. https://doi.org/10.1038/sj.npp.1301166

Wyller, V. B., Eriksen, H. R., & Malterud, K. (2009). Can sustained arousal explain the Chronic Fatigue Syndrome? *Behavioral and Brain Functions*, *5*, 10. https://doi.org/10.1186/1744-9081-5-10

Zio, E., Fan, M., Zeng, Z., & Kang, R. (2019). Application of reliability technologies in civil aviation: Lessons learnt and perspectives. *Chinese Journal of Aeronautics*, *32*(1), 143–158. https://doi.org/10.1016/j.cja.2018.05.014

Appendices

i. Description of a pilot’s duty

ii. Practical implications for reporting requirements

iii. R packages used for analysis

Appendix i. Description of a pilot’s duty

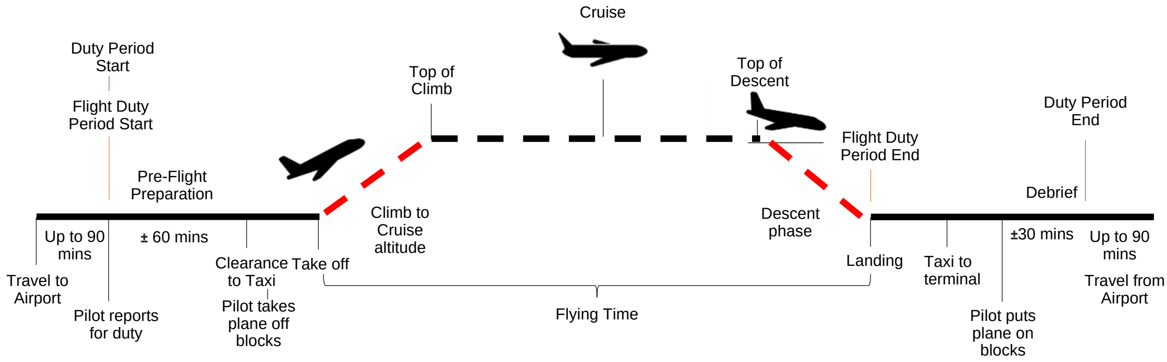


Figure 20. Illustration of a pilot’s duty. For simplicity just one flight has been depicted. For short haul operations, there may be typically 4-6 flights per duty period. The time on the ground in-between such flights is called the ‘turnaround time’.

Travel to Airport

For most airlines, the time budgeted for the commute portion of the pilot’s duty is a maximum of 90 minutes.

Pre-Flight preparation

Following reporting for duty, the pilot has a number of pre-flight duties that are undertaken during the preparation time for flight. The pilots will conduct briefings with the crew on the upcoming flight and undertake various activities including a walk around of the aircraft, (to conduct a visual inspection of the entire outside of the aircraft to check for any abnormalities). Once on the plane, pilots will also program the computers in the flight deck in accordance with their flight plan, check fuel requirements and whether fuel has been loaded, and undertake various systems checks.

Clearance to Taxi

Once the aircraft is boarded with passengers and loaded with cargo, the pilots will obtain a ground clearance from the airport control tower to taxi. The aircraft is then pushed back, ‘off blocks’ or ‘chocks’ typically with a tug. When the aircraft is clear to power under its own steam, the tug is unhooked and the aircraft will be manoeuvred to the take off runway.

Take-off

When the aircraft accelerates to a flying speed (specific to every operation), it becomes airborne and the landing gear is retracted. The actual take off speed and distance required for every flight varies due to a number of factors: pressure height, wind speed and direction, aircraft weight, air temperature, flap setting, runway gradient, clearance and operational requirements.

Climb and cruise

In the initial climb, in normal operations a power setting close to maximum thrust will be set for take off. The landing gear is retracted to help reduce drag and increase lift, as soon as the aircraft is airborne. The pilots will subsequently reduce the take off power to a lesser thrust for a climb when a safe altitude is deemed to have been reached. Roughly five to ten minutes into the climb the aircraft reaches cruising altitude, when the passenger seatbelt sign is turned off in smooth conditions. Most of the flight time is spent at cruise altitude (which will vary depending on the operation and conditions).

Descent, final approach and landing

The descent phase begins with the decrease in altitude from cruise altitude to initial approach altitude. This occurs approximately 20 minutes from the estimated time of arrival. During the approach phase, the aircraft is configured for landing through processes that act to gradually slow the aircraft down. The flaps and slats of the wing are extended to create more lift at a slower speed, which reduces the landing speed. The wheels are lowered, and the aircraft is lined up on the approach path to land. Landing is the final stage of flight where the aircraft eventually touches down on the runway tarmac. The wheel brakes are applied and the reverse engine thrust is activated, which help to slow the aircraft down to taxi speed.

Taxi to the terminal

This is where the aircraft is manoeuvred from the runway to the terminal via the taxiways to a designated arrival bay.

Post flight

This stage is where pilots undertaken post-landing checklists, procedures and fill out any debrief reports. In short haul operations, the crew will ‘turn around’ the aircraft for the following flight and follow the procedures over again for the next sector.

Appendix ii. Practical Implications for reporting requirements

Table 17. Suggested minimum reporting requirements for fatigue-related MORs

Categories	Suggested Key Data Requirements
Distal factors	Please provide numerical time and time zone data for: i. Pre-duty sleep wake history (sleep and wake times over the past 72 hours) ii. Preceding work pattern (duty start and end times and time zone information over the past 72 hours, or longer if cumulative fatigue risk issue) OR include duty roster (at least previous 7 days) within the MOR
Proximal factors	i. Please describe any 'on-the-day' deviations in work hours from planned work pattern ii. Please set out timing (hrs and phase of flight) and duration of fatigue symptoms iii. Please include self-assessment of KSS rating for sleepiness level at time of incident and the time period or phase of flight during which symptoms were experienced
Safety prompt questions	
For in-flight fatigue events	Was the other flight crew member affected by fatigue? If so over what period of flight and to what degree of severity?
For safety concerns over schedule-driven fatigue risks	Does your safety report relate to a recurring issue with a specific scheduling practice or work pattern?
For any safety concerns without proximal errors or outcomes	What factors prevented a fatigue-related serious incident safety or accident from occurring?

Appendix iii. Main R packages for analysis

Table 18. R packages

Package	Reference
afex	Henrik Singmann, Ben Bolker, Jake Westfall, Frederik Aust and Mattan S. Ben-Shachar (2022). afex: Analysis of Factorial Experiments. R package version 1.1.1. https://CRAN.R-project.org/package=afex
dplyr	Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.7. https://CRAN.R-project.org/package=dplyr
ggplot2	H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. R package version 3.3.6
graphics	R Core Development Team (2012). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
lme4	Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01. R package version 1.1.30
optimx	John C. Nash, Ravi Varadhan (2011). Unifying Optimization Algorithms to Aid Software System Users: optimx for R. Journal of Statistical Software, 43(9), 1-14. doi 10.18637/jss.v043.i09. R package version 22.4.30
Rcmdr	Fox, J., and Bouchet-Valat, M. (2022). Rcmdr: R Commander. R package version 2.8.0
stats	R Core Development Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

END