Safety and Health at Work 12 (2021) 462–470

Contents lists available at ScienceDirect

Safety and Health at Work

journal homepage: www.e-shaw.net

Original article

Proposed Data-Driven Approach for Occupational Risk Management of Aircrew Fatigue

Benjamin Zhi Qiang Seah [1], Wee Hoe Gan [2,3,4,*], Sheau Hwa Wong [1,3,5], Mei Ann Lim [6], Poh Hui Goh [6], Jarnail Singh [7], David Soo Quee Koh [4,6]

[1] *Republic of Singapore Air Force Medical Service, Singapore*
[2] *Department of Occupational and Environmental Medicine, Singapore General Hospital, Singapore*
[3] *Changi Aviation Medicine Centre, Changi General Hospital, Singapore*
[4] *Saw Swee Hock School of Public Health, National University of Singapore, Singapore*
[5] *Sleep Disorders Unit, Singapore General Hospital, Singapore*
[6] *PAPRSB Institute of Health Sciences, Universiti Brunei Darussalam, Brunei Darussalam*
[7] *Civil Aviation Medical Board, Civil Aviation Authority of Singapore, Singapore*

A R T I C L E I N F O

Article history:
Received 9 March 2021
Received in revised form
7 June 2021
Accepted 9 June 2021
Available online 17 June 2021

Keywords:
Aviation
Data driven
Fatigue risk management
Performance-based regulatory approach
Standardized protocol

A B S T R A C T

Background: Fatigue is pervasive, under-reported, and potentially deadly where flight operations are concerned. The aviation industry appears to lack a standardized, practical, and easily replicable protocol for fatigue risk assessment which can be consistently applied across operators.
Aim: Our paper sought to present a framework, supported by real-world data with subjective and objective parameters, to monitor aircrew fatigue and performance, and to determine the safe crew configuration for commercial airline operations.
Methods: Our protocol identified risk factors for fatigue-induced performance degradation as triggers for fatigue risk and performance assessment. Using both subjective and objective measurements of sleep, fatigue, and performance in the form of instruments such as the Karolinska Sleepiness Scale, Samn-Perelli Crew Status Check, Psychomotor Vigilance Task, sleep logs, and a wearable actigraph for sleep log correlation and sleep duration and quality charting, a workflow flagging fatigue-prone flight operations for risk mitigation was developed and trialed.
Results: In an operational study aimed at occupational assessment of fatigue and performance in airline pilots on a three-men crew versus a four-men crew for a long-haul flight, we affirmed the technical feasibility of our proposed framework and approach, the validity of the battery of assessment instruments, and the meaningful interpretation of fatigue and work performance indicators to enable the formulation of safe work recommendations.
Conclusion: A standardized occupational assessment protocol like ours is useful to achieve consistency and objectivity in the occupational assessment of fatigue and work performance.

1. Introduction

Fatigue in aviation is pervasive, under-reported, and potentially deadly [1,2]. Crew fatigue is 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 [3]. Fatigue results in delayed reaction time, reduced performance accuracy, lapses in sustained attention, impaired logical reasoning, compromised decision-making, inaccurate risk assessment, reduced situational awareness, and poor motivation [4]. Recognizing fatigue as a major human factors hazard with flight safety implications has prompted the International Civil Aviation Organization (ICAO) to provide implementation guidance for Fatigue Risk Management Systems (FRMS) as an alternative to the prescriptive approach.

* Corresponding author. Department of Occupational and Environmental Medicine, Singapore General Hospital, Outram Road, 169608, Singapore.
E-mail address: gan.wee.hoe@singhealth.com.sg (W.H. Gan).

2093-7911/$ – see front matter
https://doi.org/10.1016/j.shaw.2021.06.002

ICAO defines FRMS as a data-driven means of continuously monitoring and managing fatigue-related safety risks, based upon scientific principles and knowledge as well as operational experience that aims to ensure relevant personnel are performing at adequate levels of alertness [3]. The ICAO Council officially adopted the new international standards recommended for FRMS as of June 2011, which became applicable on 15 December 2011 [5]. The value proposition of FRMS in the management of fatigue-related risks specific to the aviation industry is its multi-pronged and customizable approach where varied operating circumstances (i.e., flight duration, crew composition) are concerned, whereas traditional prescriptive flight and duty time limits, stemming from historical limits on working hours during the Industrial Revolution and subsequently regulations that limited working hours in transportation sectors in the early 20th century, represent a one-size-fits-all approach, FRMS accords flexibility in operational planning and execution by adopting a performance-based regulatory approach. Operators are hence able to design bespoke solutions to meet individual operational requirements. For example, in April 2005, easyJet (a large European low cost carrier) became the first major airline to be granted derogation by the United Kingdom Civil Aviation Authority from existing limits of three consecutive early duties through an FRMS-backed proposal to institute a 5/2/5/4 roster (5 early duties, 2 days off, 5 late duties, 4 days off) as opposed to the previously practiced 6/3 roster (3 early duties, 3 late duties, 3 days off) [6].

The ICAO Standards and Recommended Practices (SARPs) describe three types of hazard identification. Predictive hazard identification examines planned work schedules, taking into account factors known to affect sleep and fatigue. Reactive hazard identification is a post-event/-incident process wherein the contribution of fatigue to safety reports filed is scrutinized. Proactive hazard identification processes (vis-à-vis predictive and reactive) examine crew fatigue and performance data during operations such that any consequent risk may be assessed and mitigated [7]. Proven tools utilized in this context include crew fatigue surveys [8] and crew performance data [9]. This framework facilitates an iterative process of fatigue onset detection that may otherwise remain unrecognized by the sole study of known risk factors or lessons from past occurrences.

FRMS regulatory requirements differ between geographical states, and its use and implementation further differ amongst regulators. The aviation industry appears to lack a standardized and easily replicable protocol for fatigue risk assessment which can be consistently applied across operators. A standardized protocol for fatigue risk assessment based on recognized risk factors and operational monitoring would achieve consistency and objectivity in flight operations decision-making and eliminate fatigue onset based on the hierarchy of controls. A large-scale European Union study of aircrew fatigue during long night and disruptive duties [10] demonstrated how protocolization of fatigue risk measurement could yield valuable and operationally-translatable conclusions across multiple airlines and expand and reinforce our knowledge in the predictive hazard identification domain. Additional benefits include a common denominator for comparison and application of physiological data obtained across operators and a larger database size enabling more robust data analysis.

Our paper sought to present a framework, supported by real-world data with subjective and objective parameters, to monitor aircrew fatigue and performance, and to determine the safe crew configuration for commercial airline operations. An operational study involving the occupational assessment of fatigue and performance in airline pilots on a three-men crew versus a four-men crew for a long-haul flight between Bandar Seri Begawan and London was conducted. This operational, real-world study affirmed the technical feasibility of our proposed framework and approach, the validity of the battery of assessment instruments, and the meaningful interpretation of fatigue and work performance indicators to enable the formulation of safe work recommendations.

2. Materials and methods

Our proposed protocol for cockpit-based aircrew directly involved in aircraft controls is illustrated (Fig. 1). Upon setting out to evaluate feasibility of a new flight route or crew composition, it is important to first and foremost analyze and identify risk factors to fatigue development and fatigue-induced performance degradation (Step 1). Such potential factors for long-haul flights include long duty hours and periods of extended wakefulness, critical high workload situations coinciding with physiological circadian nadirs as in red-eye departures, and circadian disruption from time zone crossings resulting in a poorly-rested state post-layover [11]. Of note, the fact that short-haul flights can also be fatigue-inducing must not be neglected as with the case of multisector flight schedules leading to consecutive high-intensity workload periods packed within a single workday and duty period.

To a certain extent, fatigue risk factors may be mitigated through aviation regulations and procedures (administrative controls), such as maximum crew duty period and flight time, minimum rest periods, and crew configuration. However, there is also the recognition that existing regulations cannot address all instances of fatigue and performance degradation brought about by operational complexities and individual factors. As such, effect measurements form another important aspect of risk assessment. To this end, a standardized methodology for fatigue and performance assessment should be adopted to qualitatively and quantitatively evaluate the physiological impact induced by the introduction of a known operational or crew variable (Step 2). Should one or several variables produce outcome measures that exceed predetermined safety thresholds or indicate a significant state of fatigue, mitigating measures or modifications should be instituted with re-evaluations performed until deemed satisfactory in accordance with existing safety standards to eliminate this hazard (Step 3).

Individual data from a combination of subjective sleepiness and fatigue ratings, i.e., the Karolinska Sleepiness Scale (KSS) and the Samn-Perelli Crew Status Check (SPS), and an objective performance measurement, i.e., the Psychomotor Vigilance Task (PVT), were collected as part of proactive hazard identification in our protocol, alongside ongoing sleep monitoring using both sleep logs and actigraphy. The instrument battery proposed to assess fatigue, fatigue-induced performance degradation and sleep deficit are detailed (Table 1).

A timeline detailing execution of the various hazard identification tools during the evaluation of long-haul and short-haul flights is presented (Fig. 2A and B). Preflight data were typically collected during the period of time preceding the flight under investigation from the conclusion of an aircrew's last reported flight. It is representative of the degree of restfulness and sleep recovery prior to embarking on crew duties in the next flight. Postflight data were collected from flight disembarkation up until an aircrew's next scheduled flight. It provided valuable information regarding sleep recovery after sleep restriction, its pattern and effectiveness, and the circadian disruption incurred as a result of flight schedules or time zone crossings where applicable.

For meaningful data analyses, we proposed segmenting the total crew duty period (including flight duration) into phases. This applied to both long- and short-haul flights and would allow crew fatigue and performance measurement results to be determined in the different phases over time. Individual datapoints within each phase were batched for analysis to increase overall power. Flights

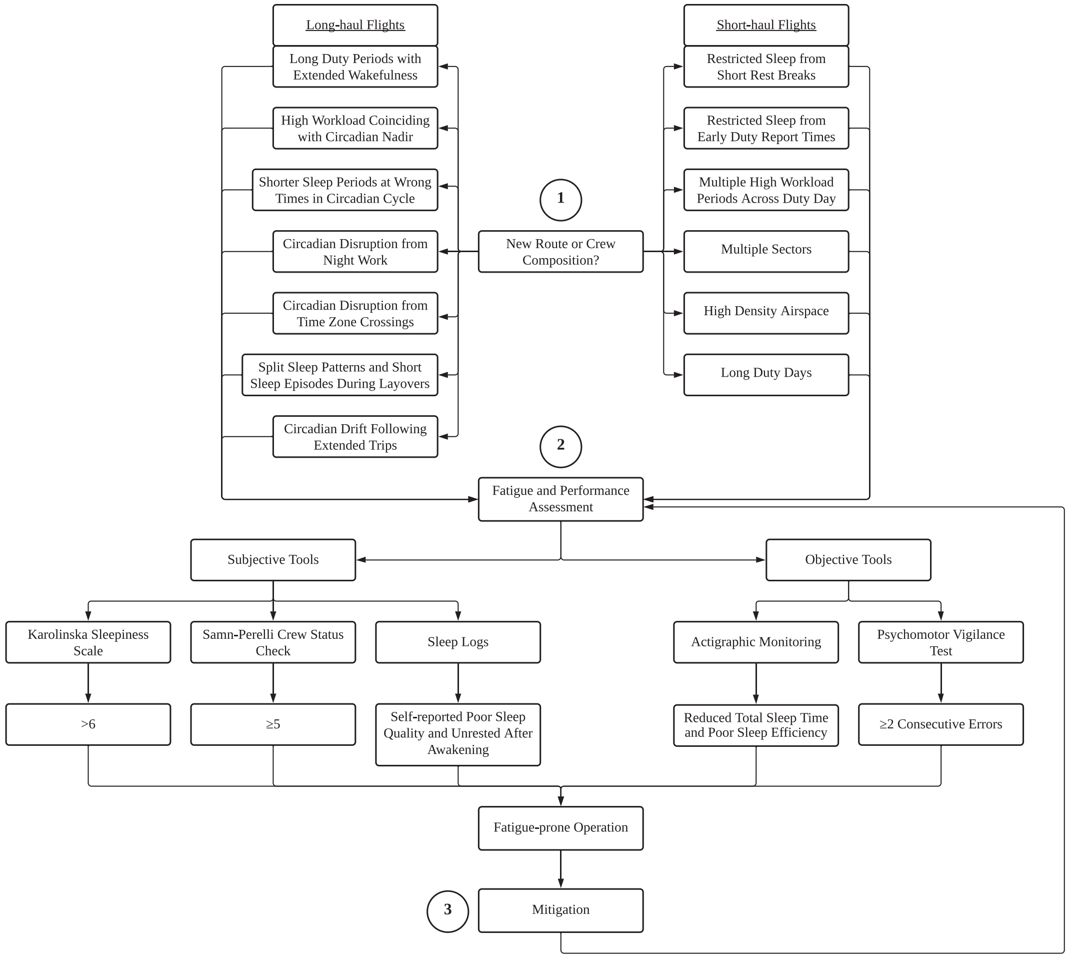

Fig. 1. Fatigue Risk Assessment Protocol.

with profiles which were longer in total crew duty period, more intense and/or operationally complex would more likely than not see significant increments in KSS, SPS, reaction time and lapses. Outcomes stemming from fatigue mitigation interventions monitored via subjective sleep logs and actigraphy could be objectively compared with either a within-subjects or between-subjects study design. A sample data presentation format with comparison of KSS means for Phases 1, 2 and 3, and between two different schedules/configurations is presented (Fig. 3).

The alert thresholds for potentially high-risk situations that might prompt a “no-go” call are listed (Table 2). These alert thresholds reflect a high probability for microsleeps, fatigue-induced performance degradation, and overall flight safety compromise. In such scenarios, handing over of flight controls to a more alert crew member and triggering of FRMS review and improvement would be mandatory.

In our operational study aimed at occupational assessment of fatigue and performance in airline pilots on a three-men crew composition for an average 14.75 hours long-haul flight from Bandar Seri Begawan to London instead of double sets of crew (i.e., four-men fully-augmented crew composition), the above protocol and instrument battery was utilized on 19 consenting aircrew out of a total of 70 (36 Captains and 34 First Officers) eligible personnel listed on the duty roster who were invited to participate in the study. Seven other aircrew who consented were unable to participate in the study due to cessation of flights as a result of the COVID-19 pandemic. These 19 aircrews included both Captains and Senior First Officers, whose cockpit job scopes were largely similar, i.e., aircraft control and navigation duties, and differed only in the level of responsibility commensurate to their corresponding rank. The participating aircrew undertook flights in both crew configurations in no particular order but were subjected to the same experimental protocol, i.e., a within-subjects study design. As such, equal weightage in data analysis and interpretation was accorded and there was no sub-analysis performed by rank. The sleep log distributed was an adapted version of a consensus sleep diary [14], which collected sleep-related information from 2 days prior to the outbound flight to 3 days after the return inbound flight, inclusive of the layover period. The PVT was self-administered by means of a PVT-192 device (Ambulatory Monitoring Inc, Ardsley, NY, USA) based on a validated 5-minute version [15]. Mean reaction time (milliseconds [ms]) and number of lapses (defined as the number of times when the reaction time was >500 ms) were measured using the PVT-192 device. An ŌURA ring (ŌURA Health Ltd, Oulu, Finland) which could continuously record actigraphic data for up to 7 days provided information on the total sleep time (TST) and sleep efficiency, i.e., TST expressed as a percentage of time in bed. A validation study evaluating ŌURA ring performance against polysomnography in measuring sleep showed 96% sensitivity in sleep detection and 90.9%, 81.3%, and 92.9% accuracy when categorizing PSG-defined TST ranges of <6 hours, 6–7 hours, >7 hours, respectively [16].

Each participating aircrew was allocated two study periods – one on a four-men crew configuration and the other on a

Table 1
Subjective and objective proactive hazard identification tools.

S/N	Tool	Description
1.	Karolinska Sleepiness Scale (KSS)	9-level validated subjective sleepiness rating scale: 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 keeping awake, fighting sleep
2.	Samn-Perelli Crew Status Check (SPS)	7-point validated subjective fatigue scale: 1 = fully alert, wide awake 2 = very lively, responsive, but not at peak 3 = okay, somewhat fresh 4 = a little tired, less than fresh 5 = moderately tired, let down 6 = extremely tired, very difficult to concentrate 7 = completely exhausted, unable to function effectively
3.	Psychomotor Vigilance Task (PVT)	A validated 5-minute version of the PVT on the PVT-192 device measures reaction time to stimulus and quantifies the number of attentional lapses, providing an objective measurement of alertness levels.
4.	Sleep Duration	Acute or cumulative sleep deficits would invariably cause fatigue and increase sleep pressure. Data on sleep duration can be collected from both self-reported sleep logs and actigraphy.
5.	Sleep Quality	Factors influencing sleep quality include: - Environmental disturbance (e.g., noise, temperature) - Luminosity (e.g., background lighting) - Personal disturbances (e.g., disruptive thoughts, illness) - Inter-personal disturbances (e.g., roommate) Alcohol, caffeinated products and medications have an impact on sleep quality: - Alcohol reduces latency to sleep onset and affects SWS and REM sleep. - Caffeinated products and medications containing caffeine interfere with sleep initiation and maintenance. Sleep quality can be assessed as a function of total sleep time and sleep efficiency. Data on sleep quality can be collected from both self-reported sleep logs and actigraphy.

three-men crew configuration, in no particular order. The study period commenced 2 days before the outbound flight, included the entire flight duration and layover period, and ended 3 days after the inbound leg. Participants wore an individually fitted ŌURA ring from the start of the study period and returned it for data download and analysis after arriving from the inbound flight. The PVT-192 device was issued prior to the outbound leg for in-flight use. Throughout both outbound and inbound flights, aircrew completed a questionnaire incorporating the KSS and SPS components and performed a 5-minute PVT at the following timepoints — preflight (before boarding aircraft), top of climb, prior to each bunk break, after each bunk break, 30 minutes before top of descent, and after landing. Participants were instructed to input sleep log entries starting from 2 days prior to the outbound flight, ending only 3 days after the inbound flight.

3. Results

Of the 19 participating aircrews, 17 were men (8 Captains and 9 Senior First Officers) and two were women (1 Captain and 1 Senior First Officer). Their mean age was 39.2 years (SD = 8.9), and the mean experience level was 17.1 years (SD = 7.5). Crew configuration ordering, demographics and data availability are described (Table 3). Results of our operational study demonstrating comparison and analysis of KSS, SPS, and PVT reaction time and lapses are presented (Fig. 4A—D). To enhance visual comparability, data from Phase 1 (pre-flight up to first 6 hours of flight time) and Phase 3 (beyond 12 hours of flight time) were presented alongside each other. Descriptive data demonstrated that the median KSS and SPS ratings, representing subjective sleepiness and fatigue levels experienced, respectively, were identical in both phases despite differences in crew configuration entailing longer flight duty times for the three-men crew set. The median Phase 1 KSS rating was 3 irrespective of crew configuration with an expected increase to 5 in Phase 3 whilst still maintaining below the predetermined threshold level of 6. Similarly, the median Phase 1 SPS rating was 3 regardless of crew configuration, with the subsequent increase to a sub-threshold median rating of 4 during Phase 3 within expectation.

There were no statistically significant differences ($p > 0.05$; unpaired t-test) within subjects in both crew compositions (the fully augmented four-men crew set being the practicing standard) across the predefined flight time intervals where objective measurements of reaction time were concerned. Mean reaction times in the four-men crew set versus the 3-men crew set were 243.2 ms (SD = 34.1) and 256.9 ms (SD = 39.0) for Phase 1, and 266.9 ms (SD = 81.6) and 272.6 ms (SD = 62.7) for Phase 3, respectively. There was no statistically significant difference in terms of mean reaction times when making comparisons between Phase 1 and Phase 3 ($p > 0.05$; unpaired T-test).

The results demonstrated that despite less time being spent by pilots resting and sleeping during bunk break (Fig. 4E and F) when operating in a three-men crew composition than that of a four-men crew composition [5.6 hours (SD = 0.89) vs. 3.8 hours (SD = 0.57); $p = 0.0001$; unpaired t-test] with comparable sleep efficiency [73.2% (SD = 21.5) vs. 77.5 (SD = 13.7); $p > 0.05$; unpaired t-test], in terms of sleepiness, crew fatigue, and performance, the 3-men crew composition is no worse off than a 4-men crew composition for long-haul flights between Bandar Seri Begawan and London. The conclusions drawn from meaningful interpretation of fatigue and work performance indicators translated into safe work

(A)

Timeline				
Pre-flight	*Outbound Flight*	*Layover*	*Inbound Flight*	*Post-flight*
Daily self-reported sleep log entries				
Actigraphy recordings				
	KSS, SPS and PVT*		KSS, SPS and PVT*	

**Performed at the following time points: pre-flight, top of climb, at each instance before and after bunk breaks, 30 min before top of descent, and after landing.*

(B)

Timeline						
Pre-flight	*Sector 1*	*Break*	*Sector 2*	*Break*	*Sector 3*	*Post-flight*
Daily self-reported sleep log entries						
Actigraphy recordings						
	KSS, SPS and PVT†		KSS, SPS and PVT†		KSS, SPS and PVT†	

†Performed at the following time points: pre-flight, top of climb, mid-flight, top of descent, and after landing

KSS: Karolinska Sleepiness Scale (KSS)

SPS: Samn-Perelli Crew Status Check (SPS)

PVT: Psychomotor Vigilance Task (PVT)

Black areas correspond to time periods where a particular subjective or objective

measurement tool was not applicable

Fig. 2. A: Application of Hazard Identification Tools During a Long-haul Return Flight. B: Application of Hazard Identification Tools During a Short-haul Flight.

recommendations during real-life flight operations decision-making where crew composition was concerned. This real-world study affirmed the technical feasibility of our proposed framework and approach and the validity of the battery of assessment instruments.

Should the results have indicated otherwise, i.e., that the three-men crew composition resulted in significantly greater fatigue and reduced performance levels, the FRMS approach mandates that the risk be mitigated. This can be accomplished through a variety of strategies including (1) schedule and roster adjustments, (2) alterations to crew composition, (3) strategic napping, and (4) pharmacological fatigue countermeasures.

4. Discussion

We described a practical and data-driven approach to performing both predictive and proactive hazard identification through the collection of validated subjective and objective fatigue and work performance indicators. More importantly, a standardized protocol with proposed threshold measures was presented to achieve consistency and objectivity in the occupational assessment of fatigue and work performance. The workflow's validity was further affirmed through an operational, real-world study conducted to examine the impact of a reduced crew composition on fatigue risk during long-haul flights.

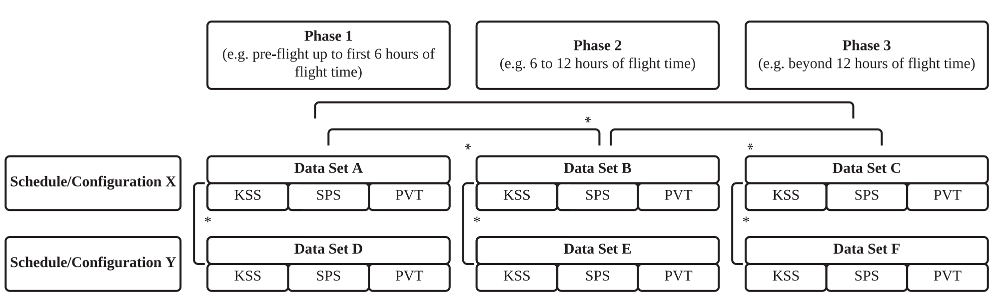

Fig. 3. Sample Data Presentation Format for Comparison of Fatigue Measurement Tools.

Table 2
Alert thresholds for potentially high-risk situations.

Hazard identification tool	Alert threshold
Karolinska Sleepiness Scale (KSS)	>6 (subjective ratings above this are associated with sleep intrusions on EEG)
Samn-Perelli Crew Status Check (SPS)	≥5 (subjective ratings above this are associated moderate fatigue with possible performance impairment)
Psychomotor Vigilance Task (PVT) (5-min version)	≥2 consecutive lapses defined as reaction time >500ms (delayed reaction times and consecutive lapses provide objective evidence of reduced alertness and performance degradation)
Sleep log	Subjectively rated poor sleep quality Subjectively-rated sensation of being not at all rested (subjective reports of poor sleep quality and an unrested state indicate unfulfilled sleep demands and sleep disruptions)
Actigraphy recording	Consistent reduction in total sleep time from baseline (recommended sleep duration for adults being ≥7 hours [12]) with poor sleep efficiency (<85% [13]) correlated with subjective sleep log reports across multiple study participants (reduced total sleep time and poor sleep efficiency provide objective evidence of a poorly-rested state)

4.1. Subjective and objective fatigue and work performance indicators

Subjective sleepiness and fatigue ratings (KSS and SPS), objective performance measurement tools (PVT), and sleep monitoring apparatuses (sleep logs, actigraphy and polysomnography) are well-described methods for proactive hazard identification.

The subjective sleepiness and fatigue self-reporting tools in our proposed protocol were selected taking into account prior validation and electrophysiological correlation studies. The relation of the KSS to electroencephalogram (EEG)/electrooculogram (EOG) indicators of sleepiness has been described to be highly significant, strongly curvilinear and consistent across individuals. High (>6) KSS values are associated particularly with impaired driving performance and sleep intrusions in the EEG [17], forming the basis of our protocol's threshold level. The KSS has been reported to possess significant correlation to EEG and behavioral indicators of sleepiness, indicating a high validity in measuring sleepiness [18]. It has also been found to be highly and significantly correlated with theta activity in both tired and rested states and with alpha activity in the rested state [19]. Notably, the KSS score, alongside EEG alpha activity, was significantly correlated with succeeding performance on the vigilance test, suggesting it could be used to predict performance errors [20]. The SPS was originally developed by the United States Air Force School of Aerospace Medicine for aircrew to reduce the time required to report fatigue data. An SPS score of 5 was estimated to reflect moderate fatigue levels with possible performance impairment for which flying duties were permissible but not recommended unless urgent [21]. Similarly, this formed the basis of our protocol's threshold level. Sleep logs provided a simple means through which sleep duration may be estimated and sleep quality and degree of restfulness rated. Subjective information and personal recounts of factors potentially contributing to poor sleep quality and inadequate rest that could have otherwise been overlooked may be obtained through sleep logs as well. It was also important to similarly capture deliberate or incidental caffeine consumption due to its impact on alertness and sleep initiation and maintenance.

The use of actigraphy for objective monitoring of sleep parameters through the use of a noninvasive accelerometer has been examined in multiple studies. These studies suggest that wrist actigraphy is useful in the estimation of TST, sleep percentage, and wake after sleep onset (WASO) [22]. When compared to the gold standard of polysomnography (PSG), actigraphy's sensitivity (actigraphy reflects sleep when PSG indicates sleep) and accuracy (total proportion of sleep captured in the actigraphy is very close or similar to that indicated by the PSG) were high, whereas specificity (actigraphy reflects an awake state when PSG indicates an awake state) was low [23]. Prior to usage, the actigraph of choice and its data interpretation algorithm must be PSG-validated for sensitivity and accuracy, with clear instructions regarding its use provided to the crew. Specifications-wise, it should ideally be lightweight, well-fitted but unobtrusive with minimal work interference, be waterproof, and possess good endurance where battery life is concerned.

The PVT allows for the objective quantification of fatigue-induced performance degradation by measuring reaction time to stimuli occurring at random intervals. The PVT demonstrates high sensitivity to sleep deprivation [24,25] and its impact on vigilant attention. Response speed and the number of lapses have high effect sizes [26] and were preferentially selected as primary outcome metrics. Feasibility of implementing briefer versions of the PVT in clinical and operational contexts has been affirmed [26]. For PVT administration to be practicable in the context of flying, the test conduct should not require a period of time longer than what is necessary to obtain essential data, especially taking into account that the test may be conducted in an operational flight. It should also have a shallow learning curve but no learning effect and be self-administered on a portable handheld device for convenience.

Our proposed use of the KSS, SPS, PVT, sleep logs, and actigraphy as tools are well-documented in the existing ICAO Fatigue Management Guide for Airline Operators. They have also been utilized and incorporated within methodologies in published fatigue risk assessment studies to critically evaluate fatigue risk and derive evidence-based recommendations [27,28]. The KSS and SPS as subjective sleepiness and fatigue ratings and the PVT as an

Table 3
Crew configuration ordering and demographics.

Crew Configuration ordering	a. 4-Men crew then 3-men crew b. 3-Men crew then 4-men crew c. 4-Men crew (3-men crew not studied) d. 4-Men crew (rejected 3-men crew)	a. 5 participants b. 1 participant c. 6 participants d. 7 participants
Gender	a. Male b. Female	a. 17 pilots (8 Captains and 9 Senior First Officers) b. 2 pilots (1 Captain and 1 Senior First Officer)
Age	Mean = 39.2 years (SD = 8.9)	Range = 27–58 years
Experience level	Mean = 17.1 years (SD = 7.5)	Range = 6.5–32.9 years

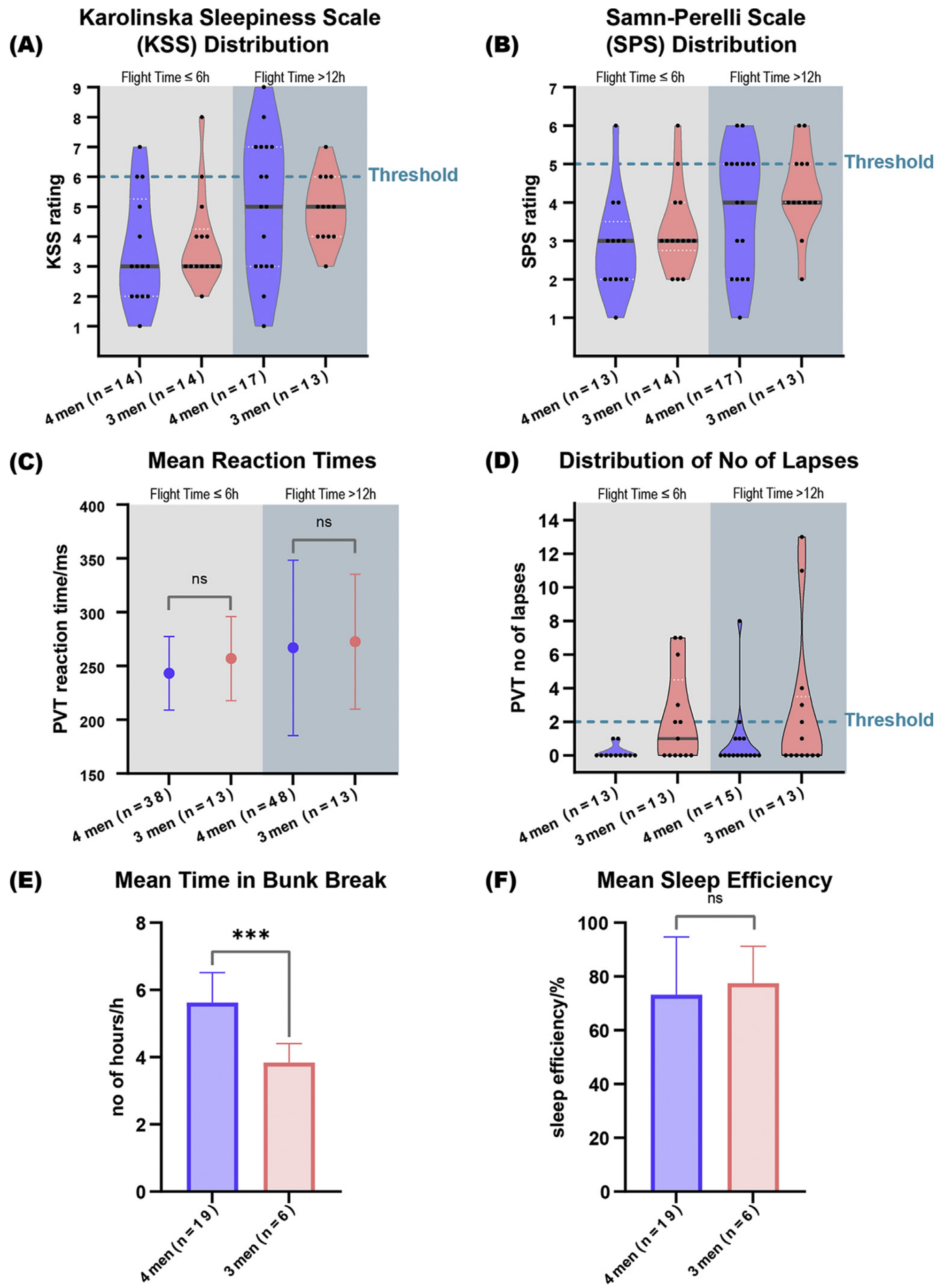

Fig. 4. Data Presentation and Analysis for Reduced versus Fully Augmented Crew Set.

objective measure of performance meet the criteria of being (1) short, intuitive, quick, and easy to complete, (2) designed to be completed at multiple time points without compromising duty performance, (3) validated, (4) predictive of objective measures of simple and complex task performance, and (5) used in aviation operations and studies with available data for comparison [7]. Pertaining to the use of actigraphy, the accepted standard for analyzing actigraphy records is for sleep diary correlation and sleep duration and quality analyses against the manually recorded sleep log. When used in a complementary fashion as in our proposed protocol, the combined data obtained from the battery of assessment instruments, analyses performed, and resultant conclusions drawn would be of high scientific credibility.

However, it must be highlighted that individually, these instruments possess inherent weaknesses. Subjective ratings are prone to individual bias and may not reliably correlate with objective measurements. Objective performance measurement tools may not accurately reflect actual operational performance and may have a motivational component to it, as in the case of the PVT. Depending on the complexity of sleep monitoring apparatuses, resource accessibility and data fidelity could affect measurement accuracy. This is a concern, in particular, for actigraphy. Although polysomnography can provide accurate information on transitions between sleep and wakefulness as well as sleep restfulness, data collection and interpretation require specialized tools and knowledge. The polysomnography set-up as described in the operational validation of the first commercial passenger ultra-long-range flight between Singapore and Los Angeles by Singapore Airlines and the Civil Aviation Authority of Singapore, even though scientifically robust, may not be practicable and easily executed by smaller operators with limited resources [7]. We feel, however, that the utilization of both subjective and objective tools exerts a complementary effect during fatigue risk evaluation and should be adopted as a best practice moving forward.

Establishing numerical limits on subjectively rated sleepiness and fatigue scales could drive negative behavior and discourage prompt fatigue-related safety reports. Although quantifying fatigue-induced degradation through reaction time and sleep duration is technologically feasible, the operational significance and thresholds for action remain debatable. A scientific and data-driven approach would serve to provide the necessary evidence base to convince all stakeholders involved.

4.2. Operational study results and FRMS applications

The results of our operational study fulfilled several key objectives. It affirmed the technical feasibility of our proposed framework and approach, the validity of the battery of assessment instruments, and the meaningful interpretation of fatigue and work performance indicators to enable the formulation of safe work recommendations. Preliminary analysis demonstrated no statistical evidence of increased fatigue risk following a reduction in crew composition from the fully augmented four-men crew set to a three-men team despite less time being available for crew rest and sleep.

Predictive hazard identification in the FRMS context aims to identify aircrew, operational and/or organizational factors which predispose individuals to fatigue-induced performance issues. Individually, these factors may not pose a significant risk to flight safety. Collectively, however, the impact of these isolated factors could be compounded, with "holes" in the Swiss cheese aligning to result in a catastrophic event. Flight scheduling, crew rostering, crew composition, on-board naps, and workload management are some of the mitigation strategies which could be implemented, with fatigue risk re-assessment using our proposed protocol performed to assess the appropriateness of interventions.

Proactive hazard identification aims to detect fatigue hazards during current operations. A combination of subjective and objective tools is optimal as fatigue is multidimensional and not an intuitively quantifiable physiological parameter. Interindividual variability in fatigue-induced performance degradation further complicates its measurement and interpretation. Subjective reports of high fatigue levels and fatigue-related performance issues, alongside self-administered prospective crew fatigue surveys provide insights on reduced alertness that might not be comprehensively captured via objective means. There is an operational aspect to such datapoints that warrant representation and acknowledgement. Objective measures of fatigue and performance are equally critical in providing the necessary reassurance where crew performance is concerned. Such measures could take the form of either reaction time testing or sleep monitoring, while noting that they are complementary rather than mutually exclusive. In the context of the PVT, two or more consecutive lapses was treated as an operationally significant threshold that would warrant follow-up actions, both immediate as well as in the longer term. The immediate actions could be in the form of handing over direct aircraft controls mid-flight whilst those in the longer term would involve schedule optimization and/or work-rest cycle adjustments. Such thresholds have a tendency to be heavily dependent on the flight operations, conditions and context at the time in question. For example, two consecutive lapses would not be acceptable during a contingency situation whereas one lapse could be attributed to a random error.

Reactive hazard identification, through root cause analysis of fatigue- and safety-related reports, aims to determine the extent to which fatigue contributed to an in-flight incident, and from thereon derive solutions to prevent a recurrence. From a holistic Safety Management System (SMS) perspective, our methodology's use case could be broadened to include reactive hazard identification, i.e., routes with high incidences of fatigue-related reports and increased frequency of deviation from mandated flight time limits, and for monitoring of established routes.

Apart from the FRMS approach, the ICAO Manual for the Oversight of Fatigue Management Approaches developed for regulators also describes the mandatory limitations and requirements to its counterpart in the form of the prescriptive approach [29]. While abiding to the definitive limits set, operators should bear in mind that both approaches should coexist and incorporating positive elements of both approaches would serve to establish a robust fatigue management framework.

Notwithstanding the advantages of the FRMS approach, critics have cited that effective implementation may be hindered by potential conflicts of interest in that fatigue specialists may undertake multiple concurrent roles (e.g. consultancy, advisory, analytical, assessment) spanning across operator and regulator domains and ultimately requires a strong pre-existing safety climate and just safety culture such that critical safety problems would not be filtered, categorized, or suppressed [30]. However, we are of the opinion that it is precisely because of these potential issues that a data-driven approach with irrefutable objectivity is required.

4.3. Future studies

Moving forward, the utilization of signal detection theory (SDT) analysis to define goals in model application, identify assumptions concerning fatigue prevalence, and perform cost–benefit analyses associated with decision outcomes [31] would further enhance the robustness and applicability of this protocol. The decision criteria can also be customized according to decision goals and personnel roles.

Although listing and singling out factors might be a crude way of performing predictive hazard identification, it can be easily performed in a checklist fashion pre-flight without incurring significant time nor effort. An iterative biomathematical model customized to an identified fatigue-inducing task and validated in an operational setting would be ideal. Our method of data collection and analysis for proactive hazard identification avoided complex experimental setups and requirements for specialized tools and personnel. There is minimal interference with flying operations, convenience, and benefits from self-administration. Most importantly, the data can be quantified, analyzed and used to further refine the fatigue risk assessment protocol described in this paper. This could form the "prototype" of a standardized FRM methodology which allows crew fatigue and performance measurements in the operational aviation setting to be performed safely, effectively, and sustainably. We advocate a data-driven means for fatigue risk analysis and management in keeping with the requirement for robust scientific evidence to drive policy implementation.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

GWH, WSH, and DK conceptualized the paper. BS, GWH, and DK designed the protocol and wrote the first draft of the manuscript. GWH, WSH, MAL, GPH, and DK undertook fieldwork. GWH, WSH, JS, and DK provided expert opinions. MAL and GPH collected data

and performed data analysis. All authors participated in the critical discussion of the manuscript and have agreed to the final version of the manuscript for publication.

References

[1] Australian Transport Safety Bureau. ATSB Transport Safety Report – fatigue experiences and culture in Australian commercial air transport pilots; 2019. https://www.atsb.gov.au/media/5775545/ar-2015-095_final_.pdf.

[2] European Cockpit Association. Barometer on pilot fatigue; 2012. https://www.eurocockpit.be/sites/default/files/2017-04/Barometer%20on%20pilot%20fatigue%2C%20ECA%202012.pdf.

[3] International Air Transport Association, International Civil Aviation Organization, International Federation of Air Line Pilots' Associations. Fatigue risk management Systems implementation Guide for operators. 1st ed.; 2011. https://www.icao.int/safety/fatiguemanagement/FRMS%20Tools/FRMS%20Implementation%20Guide%20for%20Operators%20July%202011.pdf.

[4] Caldwell JA, Caldwell JL. Fatigue in aviation: a Guide to staying awake at the stick. Burlington, VT: Ashgate Publishing Company; 2003.

[5] International Civil Aviation Organization. Formalizing new approaches to fatigue risk management Systems. ICAO J 2011;66(4). https://www.icao.int/safety/fatiguemanagement/ArticlesPublications/ICAO_Journal_Vol66_No4_FRMS.pdf.

[6] Stewart S, Holmes A, Jackson P, et al. An integrated system for managing fatigue risk within a low cost carrier. Enhancing Safety Worldwide: Proceedings of the 59th Annual IASS; Oct 2006.

[7] International Air Transport Association, International Civil Aviation Organization, International Federation of Air Line Pilots' Associations. Fatigue management Guide for airline operators. 2nd ed.; 2015. https://www.icao.int/safety/fatiguemanagement/FRMS%20Tools/FMG%20for%20Airline%20Operators%202nd%20Ed%20(Final)%20EN.pdf.

[8] Powell D, Spencer MB, Holland D, et al. Fatigue in two-pilot operations: implications for flight and duty time limitations. Aviat Space Environ Med 2008;79(11):1047–50.

[9] Balkin TJ, Bliese PD, Belenky G, et al. Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. J Sleep Res 2004;13(3):219–27.

[10] Sallinen M, van Dijk H, Aeschbach D, et al. A large-scale European union study of aircrew fatigue during long night and disruptive duties. Aerosp Med Hum Perform 2020;91(8):628–35.

[11] Gander PH, Gregory KB, Miller DL, et al. Flight crew fatigue V: long-haul air transport operations. Aviat Space Environ Med 1998;69(9 Suppl. l):B37–48.

[12] Chaput JP, Dutil C, Sampasa-Kanyinga H. Sleeping hours: what is the ideal number and how does age impact this? Nat Sci Sleep 2018;10:421–30.

[13] Reed DL, Sacco WP. Measuring sleep efficiency: what should the denominator Be? J Clin Sleep Med 2016;12(2):263–6.

[14] Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012;35(2):287–302.

[15] Roach GD, Dawson D, Lamond N. Can a shorter psychomotor vigilance task be used as a reasonable substitute for the ten-minute psychomotor vigilance task? Chronobiol Int 2006;23(6):1379–87.

[16] de Zambotti M, Rosas L, Colrain IM, et al. The sleep of the ring: comparison of the ŌURA sleep tracker against polysomnography. Behav Sleep Med 2019;17(2):124–36.

[17] Akerstedt T, Anund A, Axelsson J, et al. Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. J Sleep Res 2014;23(3):240–52.

[18] Kaida K, Takahashi M, Akerstedt T, et al. Validation of the Karolinska sleepiness scale against performance and EEG variables. Clin Neurophysiol 2006;117(7):1574–81.

[19] van den Berg J, Neely G, Nilsson L, et al. Electroencephalography and subjective ratings of sleep deprivation. Sleep Med 2005;6(3):231–40.

[20] Kaida K, Akerstedt T, Kecklund G, et al. Use of subjective and physiological indicators of sleepiness to predict performance during a vigilance task. Ind Health 2007;45(4):520–6.

[21] Samn SW, Perelli PL. Estimating aircrew fatigue: a technique with application to airlift operations. Report SAM-TR-82-21. Brooks AFB, TX: USAF School of Aerospace Medicine; 1982.

[22] Martin JL, Hakim AD. Wrist actigraphy. Chest 2011;139(6):1514–27.

[23] Marino M, Li Y, Rueschman MN, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep 2013;36(11):1747–55.

[24] Dorrian J, Rogers NL, Dinges DF, et al. Psychomotor vigilance performance: neurocognitive assay sensitive to sleep loss. Sleep deprivation: clinical issues, pharmacology and sleep loss effects. New York, NY: Marcel Dekker, Inc.; 2005. p. 39–70.

[25] Lim J, Dinges DF. Sleep deprivation and vigilant attention. Ann N Y Acad Sci 2008;1129:305–22.

[26] Basner M, Dinges DF. Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. Sleep 2011;34(5):581–91.

[27] Cosgrave J, Wu LJ, van den Berg M, et al. Sleep on long haul layovers and pilot fatigue at the start of the next duty period. Aerosp Med Hum Perform 2018;89(1):19–25.

[28] Gander PH, Signal TL, van den Berg MJ, et al. In-flight sleep, pilot fatigue and Psychomotor Vigilance Task performance on ultra-long range versus long range flights. J Sleep Res 2013;22(6):697–706.

[29] International Civil Aviation Organization. Manual for the Oversight of fatigue management approaches. 2nd ed.; 2016. https://www.icao.int/safety/fatiguemanagement/FRMS%20Tools/Doc%209966.FRMS.2016%20Edition.en.pdf.

[30] Bourgeois-Bougrine S. The illusion of aircrews' fatigue risk control. Transportation Res Interdiscip Perspect 2020;4:100104.

[31] Raslear TG, Coplen M. Fatigue models as practical tools: diagnostic accuracy and decision thresholds. Aviat Space Environ Med 2004;75(3 Suppl. l):A168–72.