European Union Aviation Safety Agency

This project has received funding from the European Union's Horizon 2020 Programme

RESEARCH PROJECT EASA.2019.C31

DELIVERABLE 2.4: SAFETY PERFORMANCE METRICS AND BENCHMARK

Effectiveness of Flight Time Limitations (FTL 2.0)

An Agency of the European Union

Disclaimer

DELIVERABLE NUMBER AND TITLE: FTL D2.4. Safety performance metrics and benchmark of the results (FINAL).
CONTRACT NUMBER: EASA.2019.C31
CONTRACTOR / AUTHOR: DLR / Daniel Aeschbach (DLR)
IPR OWNER: European Union Aviation Safety Agency
DISTRIBUTION: Public

APPROVED BY:	AUTHORS	REVIEWER	MANAGING DEPARTMENT
	D. Aeschbach (DLR)	A. Roelen (NLR)	A. Rutten (NLR)
	D. Fischer (DLR)		
	M. van Miltenburg (NLR)		
	A. van Drongelen (NLR)		

EASA:

Irina Petrova, EASA Flight Standards Directorate

Scientific Committee: Barbara Stone, Alexandra Holmes, Kristjof Tritschler

DATE: 28 March 2025

SUMMARY

Problem area

The research study FTL 2.0 aims to perform a review of the effectiveness of the current flight and duty time limitations and rest requirements contained in Commission Regulation (EU) No 965/2012. More specifically, the purpose is to add to the work performed during the first phase of the "Effectiveness of Flight Time Limitation" evaluation (MOVE/C2/2016-360).

This research study includes an assessment of the impact on aircrew alertness of the following aircrew duty periods:

a) Duties of more than 13 hours at the most favourable time of the day;

b) Duties of more than 11 hours for crew members in an unknown state of acclimatisation;

c) Duties including a high level of sectors (more than 6); and

d) On-call duties such as standby or reserve followed by flight duties, specifically focussing on 'other than airport standby'.

It also comprises an assessment of the impact on aircrew alertness of controlled rest: this includes an analysis of the conditions and circumstances under which aircrew members take controlled rest.

Description of work

This deliverable includes a comparison of the key findings from Task 2.3 with predictions from biomathematical modeling, as well as a comparison with the published literature, using the systematic literature review conducted in Task 1.1 and additional literature published since then.

Results and Application

Comparison of the actual fatigue levels with the Boeing Alertness Model (BAM) predictions was successfully undertaken. Analysis of BAM-derived fatigue scores mostly matched the main conclusions of Task 2.3. Fatigue levels at last TOD of FDPs starting at the most favourable time of day (i.e., 06:00 – 13:29) were increased for FDP durations ≥11 h, with similar fatigue levels between FDPs 11-13h and FDPs >13h. Analysis of FDPs flown during the night in an unknown state of acclimatisation showed elevated levels of fatigue for FDP durations ≥ 9 h. Fatigue continued to increase beyond 6 sectors, with the occurrence probability of high fatigue suggesting an accelerated pace beyond this threshold.

The previous scientific literature review conducted in Task 1.1 identified a lack of studies addressing the effects of long FDPs starting during the most favourable time of day, long FDPs flown in unknown state of acclimatisation, and a high number of sectors on crew members' fatigue. No additional relevant studies were identified since then, and the current data collection and analysis effort represents an important contribution to fill this gap.

CONTENTS

ABBREVIATIONS

ACRONYM	DESCRIPTION
BAM	Boeing Alertness Model
BMM	Biomathematical model
COVID-19	Coronavirus disease 2019
CR	Controlled Rest
DLR	German Aerospace Centre
EASA	European Aviation Safety Agency
EU	European Union
FDP	Flight Duty Period
FTL	Flight Time Limitation
HEMS	Helicopter Emergency Medical Services
KSS	Karolinska Sleepiness Scale
NLR	Royal NLR - Netherlands Aerospace Centre
ORO	Organisation Requirements for Air Operations
PVT	Psychomotor Vigilance Test
SAFTE	Sleep, Activity, Fatigue, and Task Effectiveness model
SP	Samn-Perelli Fatigue Scale
ToD	Top-of-Descent
WOCL	Window Of Circadian Low

1. Introduction

1.1 Project description

The objective of the second phase of the study on Effectiveness of Flight Time Limitations (EASA.2019.C31) (hereinafter referred as FTL 2.0), is to perform an assessment of the impact on aircrew alertness of the following aircrew duty periods:

FDP1: Duties of more than 13 hours at the most favourable time of the day;
FDP3: Duties of more than 11 hours for crew members in an unknown state of acclimatisation;
FDP4: Duties including a high level of sectors (more than 6); and
FDP5: On-call duties such as standby or reserve followed by flight duties.
In addition, it includes an analysis of the conditions and circumstances under which aircrew members take Controlled Rest (CR).

Two series of deliverables are provided in FTL 2.0:

- D1: deliverables on the work performed in Task 1.1 on the definition of baseline and Task 1.2 on the definition of the target crew population;
- D2: deliverables on the work performed in Tasks 2.1 on the definition of scope and process for the data collection; in Task 2.2 on the data repository; in Tasks 2.3, 2.4, and 2.5 on the data analyses and benchmark against other reference sources and the synopsis of the results of the previous and current contract including a list of generally applicable performance metrics; as well as in Tasks 2.5, 2.6 and 2.7 on the analysis of effectiveness of prescriptive FTL and the conclusions and recommendations.

1.2 This deliverable

The main objective of Task 2.4 was to benchmark the collected fatigue data on the basis of two types of comparisons:

- A comparison of the actual fatigue ratings from Task 2.3 with biomathematical model (BMM) predictions based on the individual duty rosters and sleep data; and
- A comparison of the findings from Task 2.3 with the established literature, using the systematic literature review conducted in Task 1.1 as well as additional literature published since then.

Importantly, this task is not aimed at validating biomathematical models but aims to compare the overall fatigue trends based on the regression results from Task 2.3 with those predicted from the Boeing Alertness Model (BAM), a well-known and widely used biomathematical model, as a way to benchmark the actual findings against established tools of fatigue risk management.

2. Methods

The FDPs 1 (“Duties of more than 13 hours at the most favourable time of the day”), 3 (“Duties of more than 11 hours for crew members in an unknown state of acclimatisation”), and 4 (“Duties including a high level (>6) of sectors”) were selected for comparison with predictions from biomathematical models (BMMs). These three FDPs have a clear regulatory reference (i.e., approaching a regulatory threshold), in which context BMMs are frequently applied tool to check crew rosters against regulatory limits.

2.1 General approach

The main findings from Task 2.3, based on actual fatigue, were compared with predictions from biomathematical fatigue modelling. The model benefited from two data inputs: individual duty roster data and sleep-wake data.

To this end, we calculated the regression models established in Task 2.3 for each FDP, using BAM-derived fatigue scores as the outcome variable. Based on the regression results, BAM-derived fatigue was then predicted and plotted for the main aspect of each FDP (e.g., duration, number of sectors). For example, in FDP 1, the regression results for BAM-derived fatigue are presented for 6 – <11h vs. 11 – 13h vs. >13h FDPs. This allows us to directly compare fatigue trends in the real-life dataset with the biomathematically modelled fatigue. The key figures from Task 2.3 are presented to illustrate the overlap (or difference) between actual KSS vs. modelled fatigue ratings, thereby facilitating cross-comparison between deliverable reports D2.3 and D2.4. All analyses are based on the same samples and applying the same selection criteria for each FDP, as presented in Task 2.3 (for details, please refer to deliverable report D2.3).

In the Discussion, the findings are put into a wider context, using published literature as summarised in Task 1.1 (Definition of baseline), as well as additional studies that have been published since then (covering 01/2022 – 01/2025).

2.2 Literature search update

A search in PubMed was conducted on January 10th 2025, using the same search terms as for the search strategy in Task 1.1 (excluding the outcome terms “measurement” and “data collection”) but limiting the publication period to 2022/01/23 – 2025/01/10, to check for relevant publications since the last search was conducted.

2.3 Biomathematical modelling

Two well-known and widely used BMMs were applied: the Boeing Alertness Model (BAM) to predict alertness and the Sleep, Activity, Fatigue, and Task Effectiveness model (SAFTE) to predict performance effectiveness. For BAM, we used version 3.3.1 (the link to the technical fact sheet is provided as part of the Bibliography). For SAFTE, we used the open source SAFTE package from GitHub (link to the published implementation is provided as part of the Bibliography). The same data was inputted into both models, i.e. flown duty roster data and sleep-wake data. In rare cases (<1%), where there was no sleep-wake data, BAM predictions of sleep and wakefulness were used to ‘fill the gaps’, which were then used in both models. This is different to how operators usually use BMMs, in which they do not input individual sleep-wake data but rely on the model predictions of sleep and wakefulness. Since Task 2.4 does not aim to validate sleep-wake predictions of BMMs nor the habitual use of BMMs as part of airlines’ FRM systems, we used individual sleep-wake data as input to the models. Using individual data generally increases agreement between actual and modelled fatigue.

The outputs of both models are well-correlated (r = 0.91; 7,102 observations using the full dataset from Task 2.3) illustrated below in Fig. 2.3-1. We only report BAM results, since BAM outputs KSS values allowing for a direct comparison between the KSS model predictions and actual KSS ratings. SAFTE offers a conversion from performance effectiveness scores to KSS values, which introduces an additional source of variation that may impact the (non-)agreement between fatigue predicted by a model and actual fatigue. This additional source of variation can complicate interpretation (e.g., in case of non-agreement between actual and modelled fatigue, the deviation could stem from the transformation itself).

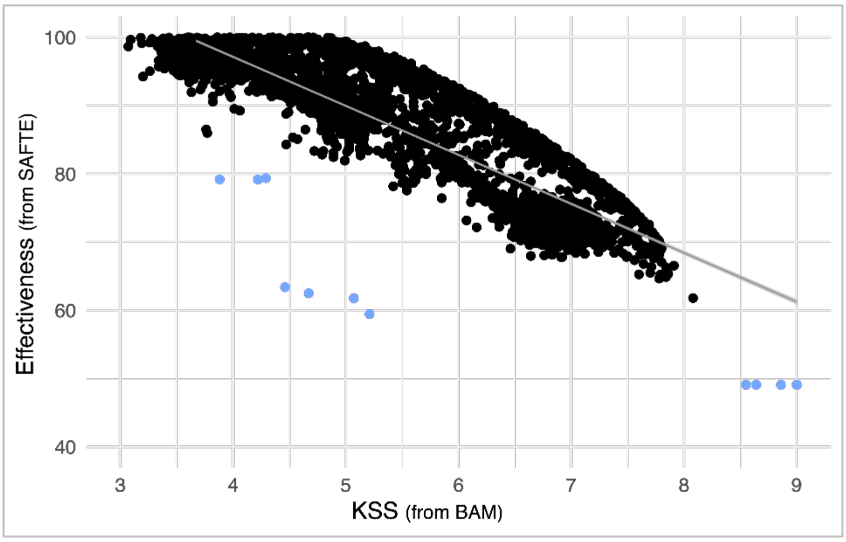

Figure 2.3-1. Relationship between SAFTE vs. BAM outputs. Higher effectiveness scores are associated with lower fatigue values on the KSS. Marked in light blue are outliers that are most likely due to translating the actual dataset into a data format that is 'readable' by the models, which can cause e.g., date issues that result in >24-h long flight activities. These outliers were excluded from the analyses.

3. Results

3.1 Literature review update

The search yielded 625 hits, of which 11 abstracts were identified as potentially relevant. None of the studies were included after full text screening, due to the following reasons:

A series of five papers was published on fatigue and cognitive performance of crew members flying under Chinese standard vs. Chinese COVID-19 outbreak response exemption policy including long-haul and long-duration duties. Three of them did not report sufficient details of the rosters to derive state of acclimatisation or the impact of long-duration FDPs starting at the most favourable time of day (Li et al., 2022, 2023; Sun et al., 2022), and a fourth focused on constructing a biomathematical model and did not include empirical fatigue scores (Sun & Sun, 2022). The fifth paper reported FDPs of more than 11 hours flying eastwards (Asia – North America) and westwards (Asia – Europe) (Sun et al., 2024). Crew members were in a known state of acclimatisation on both, outbound and inbound flights after short layovers. Importantly, crew size was augmented to 8 members according to the COVID-19 exemption policy, and reported fatigue did not exceed level 5 on the KSS ('neither alert nor sleepy') on any flight. Due to its crew size, the study was not included after full text screening.

One study examined jet-lag after various long-haul flights across time zones but did not include fatigue measures (Ruscitto et al., 2023).

Another study examined the effect of circadian disruption on fatigue and cognitive performance after long-haul flights across time zones, but did not report sufficient details to determine crew members' state of acclimatisation during the flights (Yang et al., 2024).

A 2023 study assessed fatigue and cognitive performance on two different rosters involving (ultra-)long-haul flights in military personnel, operated by augmented crews (Gläsener et al., 2023). Of a total of seven legs, one flight was longer than 13 hours and departed at 10:58 (i.e., within the most favourable time of day window, according to the definition used in FDP1). However, no control FDP of <13h departing between 6:00 – 13:29 and flown in a known state of acclimatisation was available in the study. The last legs of the rosters were apparently flown in an unknown state of acclimatisation but fatigue levels were analysed within each leg and not between flight legs. Thus, fatigue levels were shown to increase with time into FDP but not compared between FDPs flown in a known vs. unknown state.

One study examined a 3-leg roster with ultra-long range flights on the first and last leg (Sammito et al., 2022), comparing two FDPs with similar duration in different states of acclimatisation, but did not include two FDPs in an unknown-state of different durations.

Another study investigated 15-h duties starting at 6:00 in Helicopter Emergency Medical Service (HEMS) pilots but did not include fatigue or cognitive performance measures (Flaa et al., 2022).

Finally, one study examined a large number of FDPs starting at different times of day, including mid-morning start times between 7:00 – 10:59 but the focus was on comparing duty timing not durations, and FDP durations were far below 13 hours (7.30 $\pm$ 1.63 h) (Arsintescu et al., 2022).

3.2 Duties of more than 13 hours at the most favourable time of the day (FDP1)

The study of FDP1 examined fatigue levels of FDPs longer vs. shorter than 13 hours starting at the most favourable time of the day, (i.e., FDP start time between 6:00 and 13:29). The statistical multi-variable regression model applied included age, gender, crew category, sectors, time awake, time of day, and 24h sleep duration to predict: (i) actual KSS ratings, as reported by the participants; and (ii) biomathematically modelled KSS ratings, as derived from the BAM.

Figure 3.2-1 shows the continuous KSS predictions at last ToD of FDPs with a start time between 6:00 and 13:29, for the actual KSS data vs. BAM-derived KSS data. Across FDP durations, the regression models show good agreement between BAM-derived fatigue and actual fatigue, as indicated by the overlapping 95% confidence intervals. Both predictions show increasing fatigue levels with longer FDP durations, staying below the threshold of high fatigue (KSS ≥7) at last ToD.

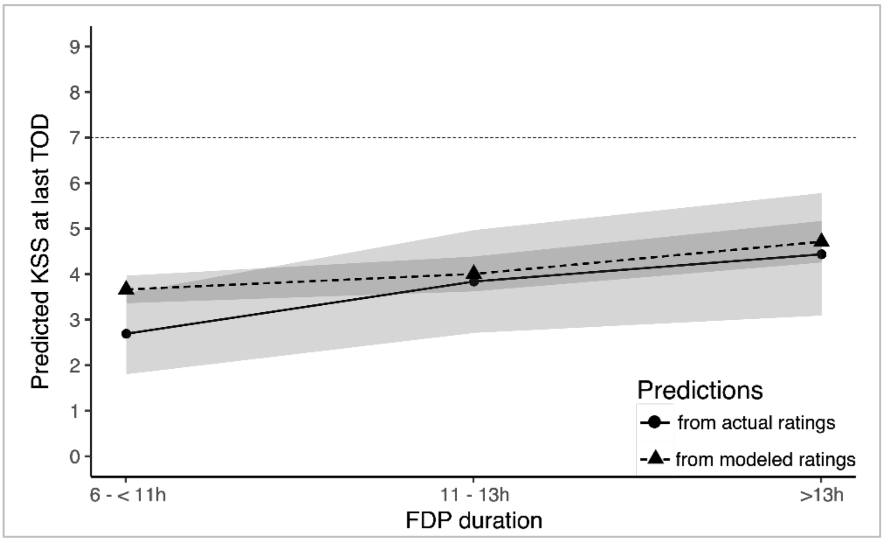

Figure 3.2-1. Predicted continuous KSS values at last ToD of FDPs with a start time between 6:00 and 13:29, based on actual KSS ratings vs. BAM-derived ratings. Grey shading represents 95% confidence intervals. N = 86 participants, 148 FDPs.

Figure 3.4-12 shows the predicted probability of high fatigue (KSS ≥7) at last ToD of FDPs with a start time between 6:00 and 13:29, for actual KSS vs. BAM-derived KSS. The regression results show good agreement between BAM-derived fatigue and actual fatigue, as indicated by overlapping 95% confidence intervals. The agreement is highest for FDP durations of ≥11 hours. For durations <11 hours, BAM appears to slightly over-estimate the probability of high fatigue at last ToD.

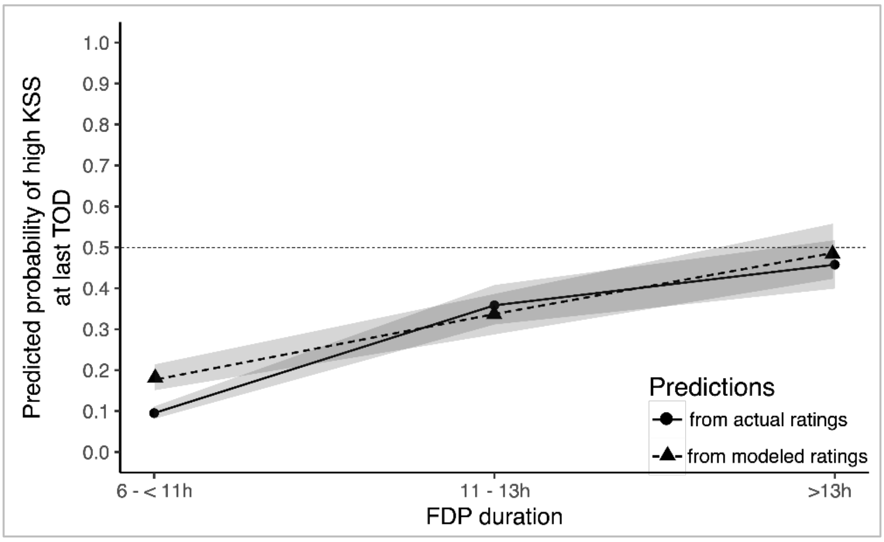

Figure 3.2-2. Regression results showing the predicted probability of high fatigue (KSS ≥ 7) at last ToD of FDPs with a start time between 6:00 and 13:29, based on actual data vs. BAM-derived ratings. Grey shading represents 95% confidence intervals. N = 86 participants, 148 FDPs.

3.3 Duties of more than 11 hours for crew members in an unknown state of acclimatisation (FDP3)

The study of FDP3 examined fatigue levels of FDPs flown in an unknown state of acclimatisation with durations of longer vs. shorter than 11 hours. The statistical multi-variable regression model applied included age, gender, crew category, sectors, time awake, time of day, 24h sleep duration, rotation/direction, time zones crossed, and layover length to predict: (i) actual KSS ratings, as reported by the participants; and (ii) biomathematically modelled KSS ratings, as derived from the BAM.

Figure 3.3-1 shows the continuous KSS predictions at last ToD of X-state FDPs for actual KSS data vs. BAM-derived KSS data. For FDP durations of less than 9 hours, the BAM model appears to overestimate fatigue of FDPs flown by crew members in an unknown state of acclimatisation. For FDP durations from 9 to more than 11 hours, the regression results show good agreement between BAM-derived fatigue and actual fatigue levels, as indicated by overlapping 95% confidence intervals. Both predictions show increasing fatigue levels with longer FDP durations, approaching or – in the case of the model predictions – reaching the threshold of high fatigue (KSS ≥7) at last ToD for FDPs longer than 11 hours flown in an unknown state of acclimatisation.

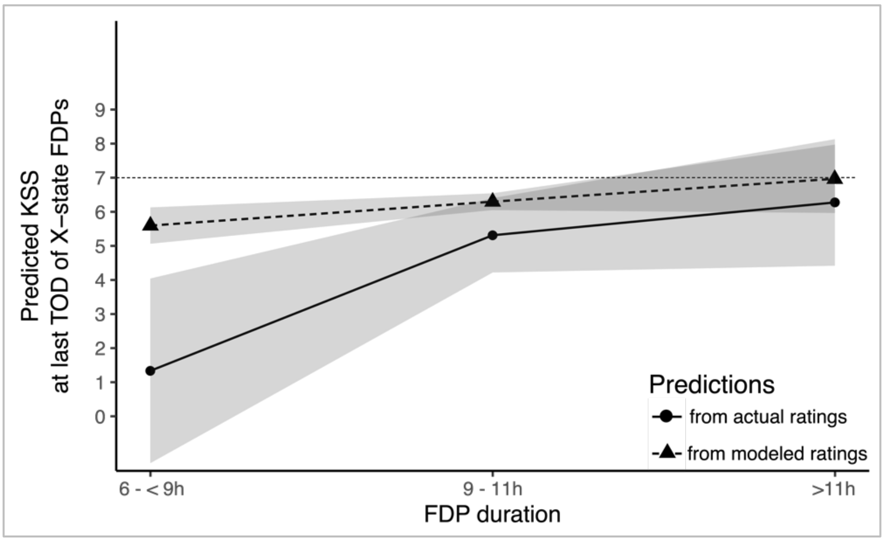

Figure 3.3-1. Regression results showing the predicted continuous KSS values at last ToD of X-state FDPs, based on actual data vs. BAM-derived ratings. Grey shading represents 95% confidence intervals. N = 21 participants, 21 FDPs.

Figure 3.4-12 shows the predicted probability of high fatigue (KSS ≥7) at last ToD of X-state FDPs for actual KSS vs. BAM-derived KSS. Across FDP durations, the regression results show good agreement between BAM-derived fatigue and actual fatigue, as indicated by overlapping 95% confidence intervals. Both approaches predict high fatigue to be more likely with increasing FDP durations, crossing the predicted probability of 0.5 for durations of ≥9 hours.

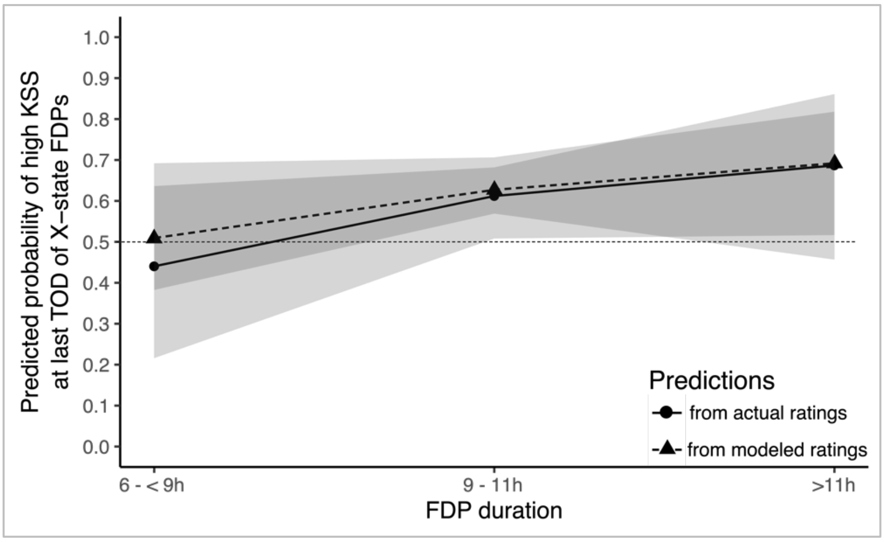

Figure 3.3-2. Regression results showing the predicted probability of high fatigue (KSS ≥ 7) at last ToD of X-state FDPs, based on actual data vs. BAM-derived ratings. Grey shading represents 95% confidence intervals. N = 21 participants, 21 FDPs.

3.4 Duties including a high level (>6) of sectors (FDP4)

The study of FDP4 examined fatigue levels of FDPs with more vs. fewer than 6 sectors. The statistical multi-variable regression model included age, gender, crew category, sectors, time awake, time of day, and 24h sleep duration to predict: (i) actual KSS ratings, as reported by the participants; and (ii) BAM-derived KSS ratings.

Figure 3.4-1 shows the continuous KSS predictions at last sector of multiple-sectors FDPs for actual KSS vs. BAM-derived KSS. There was good correspondence between the results of the two calculations, both reflecting an increasing trend of fatigue across increasing number of sectors. While the fatigue values tended to diverge slightly for higher number of sectors, the difference between actual fatigue and BAM-derived fatigue remained below one KSS point, even for 9 sectors (BAM-derived KSS 6.5 vs. actual KSS 5.9).

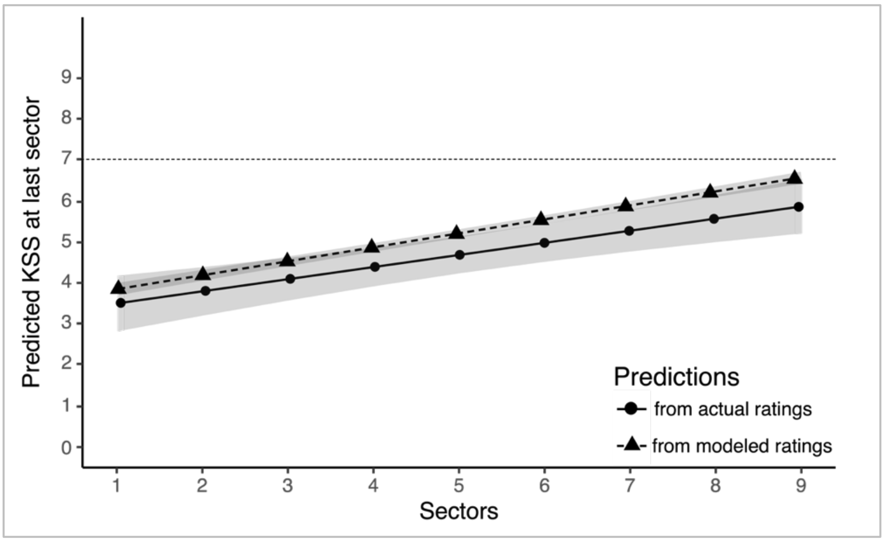

Figure 3.4-1. Regression results showing the predicted continuous KSS values at last sector, based on empirically-assessed data vs. biomathematically modelled ratings. Grey shading represents 95% confidence intervals. N = 77 participants, 362 FDPs.

Figure 3.4-2 shows the predicted probability of high fatigue (KSS ≥7) at the last sector of multiple-sectors FDPs for actual KSS vs. BAM-derived KSS. Overlapping 95%-confidence intervals at each sector indicate a fairly good correspondence between actual and BAM-derived probabilities of high fatigue occurrence. Specifically, both calculations predict an increase in the occurrence probability with number of sectors. An apparent underestimation of probabilities for the mid-range number of sectors (3-7) may reflect the fact that the effect of number of sectors in the BAM is relatively weak, compared to the effects that time awake, time of day, and prior sleep debt have on fatigue. Despite this fact, BAM-derived fatigue and actual fatigue show still good agreement for continuous KSS values (shown in Fig. 3.4-1).

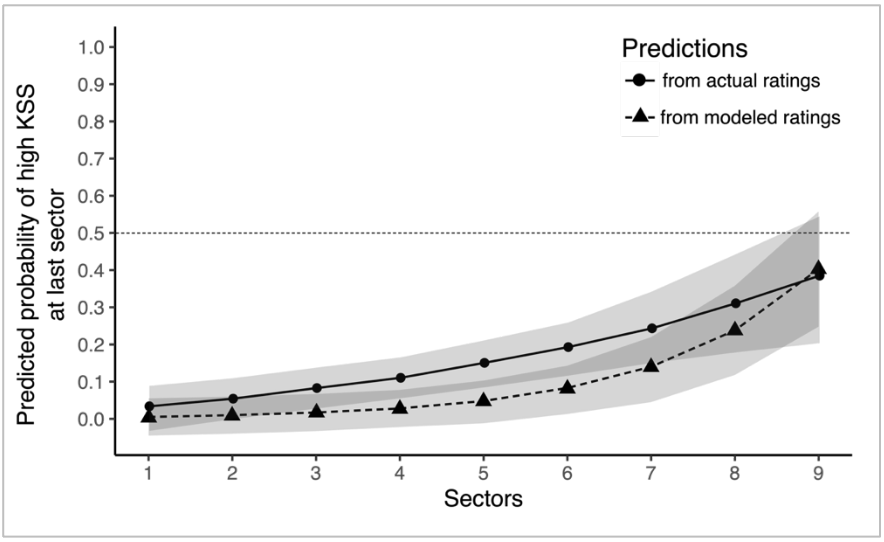

Figure 3.4-2. Regression results showing the predicted probability of high fatigue (KSS ≥ 7) at last sector, based on empirically-assessed data vs. biomathematically modelled ratings. Grey shading represents 95% confidence intervals. N = 77 participants, 362 FDPs.

4. Discussion

A literature review was conducted in Task 1.1 of this project to establish a baseline, assessing the published evidence regarding the FDPs of interest. Based on that synthesis and including new findings published since the last search that was conducted in January 2022, this chapter discusses the results from Task 2.3 and how they compare with biomathematically modelled fatigue predictions.

In task 1.1, 46 studies were included to define a baseline, ranked from low to medium and high regarding their relevance for the project. A high relevance score was assigned when the study included: the FDP of interest, repeated measures, validated measures (i.e., KSS, PVT, SP), and FDP-relevant time points (i.e., on the inbound flight). The collected state of the art was subsequently mapped using the extracted study data and synthesising the findings of the medium and high relevance studies. An update of the search for the period between January 2022 – January 2025 yielded no additional relevant studies.

4.1 Duties of more than 13 hours at the most favourable time of the day (FDP1)

The analyses in Task 2.3 regarding duties of more than 13 hours at the most favourable time of day (i.e., start time between 6:00 and 13:29) showed that fatigue levels increased with increasing FDP durations, but that most of the increase took place before the start of the 14th hour, i.e., for FDP durations of ≥ 11 hours. The BAM-derived KSS showed a matching trend, indicated by overlapping 95% confidence intervals. Both approaches, predicting fatigue levels based on actual vs. BAM-derived KSS ratings, show FDP duration to be a key determining factor of fatigue at last ToD of FDPs flown at the most favourable time of the day. Importantly, for all three FDP studies (FDP1, FDP3, and FDP4), we used individual sleep-wake schedules as an input to the BAM, which is different from how most operators use BMMs, where they predict fatigue based on roster data only. Using individual sleep-wake data generally increases the overlap between model-derived fatigue and actual fatigue. Note that the purpose of Task 2.4 was not to validate BAM predictions of sleep and wakefulness nor to validate the use of BMMs as part of airlines' FRM systems but to use a comparison of actual and modelled fatigue to further examine fatigue trends associated with specific FDP characteristics, such as duration and number of sectors.

The baseline established in Task 1.1 identified eight studies of low relevance, but yielded no medium to high relevance studies for this FDP, revealing a lack of research, particularly in 2-pilot crews, since none of the studies examined fatigue of non-augmented crews (Devine et al., 2021; Gander et al., 2013, 2014; Goode, 2003; Lamond et al., 2006; O'Hagan et al., 2020; Signal et al., 2014; Wu 2003). In addition, the long FDPs in most studies encroached (part of) the night (i.e., the least favourable time of day), limiting the transferability of the results to FDPs at other times of day. Overall, the findings of the low-relevance studies suggested that FDP duration and time of day interact on fatigue and cognitive performance, indicating that assessments of long FDP durations should take into account the timing of the duty period (i.e., extent of encroachment of the WOCL). Of the FDPs included in the analysis in Task 2.3, 67% of FDPs >13h and 56% of FDPs 11-13h fell into the category of late finish FDPs (FDP end time between 23:00 and 1:59). In addition, 9% of the FDPs >13h were night FDPs and 24% were non-disruptive schedules ("day-FDPs"), while in the category of shortest FDPs, 80% of the FDPs were non-disruptive schedules. A possible explanation for the finding that fatigue levels were similar between FDP >13h and FPD 11-13h is that the longest FDPs averaged 13:31 hours and were therefore on average only 1.5 h longer than the FDPs of 11-13h. Taking into account the previous studies, it can be assumed that even longer FDPs would further increase fatigue levels, especially because such long FDPs would most likely encroach the WOCL (02:00h – 05:59h), exacerbating fatigue via circadian effects. Hence, fatigue during long FDPs starting at the most favourable time of day can be further aggravated by time-of-day effects, when the duty period is long enough to extend into the least favourable time of day. The present study on long FDPs that start between 6:00 and 13:29 is therefore a highly needed addition to the existing body of literature, underscoring the importance of FDP duration for fatigue.

4.2 Duties of more than 11 hours for crew members in an unknown state of acclimatisation (FDP3)

The analyses in Task 2.3 regarding duties of more than 11 hours for crew members in an unknown state of acclimatisation concluded that for X-state FDPs, fatigue was significantly increased for FDP durations of 9 hours or more compared to FDP durations of less than 9 hours. The sample size was limited to 21 X-state FDPs mostly operated overnight and arriving in Europe early the next morning, while daytime X-state FDPs were mostly lacking. A similar trend was found for BAM-derived KSS, showing similar mean values and overlapping 95% CIs for X-state FDPs with a duration of ≥9 hours. For the lowest category of FDP durations in the analysis (i.e., 6 – 8.99h), the BAM over-estimated fatigue, predicting a ~4 point higher KSS mean value than actually observed.

In the literature review in Task 1.1 eight studies of medium (n=3) and high relevance (n=5) (Petrilli et al., 2006; Powell et al., 2010; Roach et al., 2012; Sallinen et al., 2021; Samel et al., 1987, 1997; van den Berg et al., 2015; Wegmann et al., 1986) were identified. None of these studies explicitly aimed to examine FDPs of different durations flown in an unknown state of acclimatisation, but reported sufficient details to approximate the state of acclimatisation, based on times zones crossed, FDP/flight times and layover length. Only one study allowed to directly compare the exposure FDP (i.e., >11h X-state FDP) with the primary control FDP (<11h X-state FDP), yet was likely confounded by time into trip, with the <11h X-state FDP being the last leg in the 4-leg trip and showing the highest fatigue scores (Petrilli et al., 2006).

In conclusion, there was a great need for studies to examine the effect of FDP duration on fatigue of crew members in an unknown state of acclimatisation. The present study comparing fatigue during X-state FDPs of different durations is therefore a much needed effort to address this gap. However, the effort remains incomplete as the majority of the 21 X-state FDPs were duties overnight.

4.3 Duties including a high level (>6) of sectors (FDP4)

With regards to FDP4, the data gathered in this project showed that flying >6 sectors during an FDP is associated with more fatigue than with fewer sectors, and that fatigue increases gradually across sectors, even beyond 6 sectors. BAM showed good agreement for continuous fatigue ratings but under-estimated the probability of high fatigue occurrences associated with the number of sectors. Existing data sets with a high number of sectors are scarce, and thus the use of biomathematical model predictions represented an important benchmarking approach.

The literature review in Task 1.1 identified 7 studies of medium (n=4) and high relevance (n=3). Studies examing FDPs with fewer than 6 sectors identified number of sectors, FDP duration and subjective workload as factors contributing to fatigue (Aeschbach et al., 2017; Powell et al., 2007; Vejvoda et al., 2014). A study by Goffeng et al. (2019) examined the effects of 10-20 sectors over a period of 4 consecutive duty days, and reported an increase of reaction time with every additional sector both in pilots and cabin crew members. Cognitive slowing is known to be an indicator of fatigue (Brezonakova, 2017). In the most recent study, Åkerstedt et al. (2021) found that the number of sectors predicted fatigue in a simple regression analysis, but the significant effect disappeared in the multi-variable analysis, mainly due to the colinear influence of FDP duration. That study examined a range of 1-8 sectors and thus was the only published one that examined fatigue for FDPs with more than 6 sectors. Although number of sectors, FDP duration and subjective workload often increase together, their influences on fatigue appear to be separate and additive. This was reported before in a simulator study (Honn et al., 2016) in which the number of sectors (1 vs. 5) was varied while duty duration was kept constant, as well as in a study with pilots during short-haul operations, which uncovered additive effects of number of sectors (1-4) and time awake. Similarly, the current sub-study identified number of sectors (1-9) and FDP duration as additive predictors of the level of fatigue, although for the occurrence probability of high fatigue only FDP duration was found to reach significance. In both cases, however, fatigue during multiple-sector duties depended on sleep duration in the preceding 24 hours, reaffirming the importance of adequate amounts of sleep within this time period. This finding is consistent with previous work identifying sleep within the prior 24 hours as a mitigating factor for increased fatigue during disruptive schedules (Sallinen et al., 2020).

5. Conclusion

The actual fatigue levels seen in the data collection effort were successfully benchmarked with the fatigue levels reported in the literature, and predicted by a biomathematical model. The previous literature review identified a great need for studies addressing the effects of long FDPs starting during the most favourable time of day, long FDPs flown in unknown state of acclimatisation, and a high number of sectors on crew members' fatigue. No additional relevant studies were identified since then, and the current data collection and analysis effort represents an important and much needed addition to fill this gap.

FDP 1. Fatigue levels at last ToD of FDPs starting at the most favourable time of day (i.e., 6:00 – 13:29) were increased for FDP durations ≥11h, with similar fatigue levels for FDPs 11-13h and FDPs > 13h. The difference between FDPs <11h and FDPs ≥11h was smaller when based on biomathematical fatigue predictions.

FDP 3. Comparison between actual and BAM-derived fatigue was limited to 21 X-state FDPs, the majority of which were flown overnight. Analysis of these unknown-state FDPs suggested significantly elevated actual levels of fatigue for FDP durations ≥9 hours. Comparison with BAM-derived fatigue, based on the same duty rosters, over-estimated fatigue for FDPs of less than 9 hours, but showed similar levels for FDP durations of ≥9 hours.

FDP 4. Regarding the analysis of multiple-sectors FDPs, the continued increase of fatigue beyond 6 sectors – and the occurrence probability of high fatigue even suggest an accelerated increase beyond this mark – is an important finding. Current flight time limitations include a stepwise malus for increasing number of sectors beyond 2 within an FDP. The current data provide justification for such a provision.

6. Bibliography

Aeschbach, D., Vejvoda, M., Mendolia, F., Tritschler, K. (2017). Additive Effects of the Number of completed flights and time awake on fatigue in short-haul airline pilots. *Sleep,* 40 (Suppl.), A58.

Åkerstedt, T., Klemets, T., Karlsson, D., Habel, H., Widman, L., & Sallinen, M. (2021). Acute and cumulative effects of scheduling on aircrew fatigue in ultra-short-haul operations. *Journal of Sleep Research*, *30*(5), e13305. https://doi.org/10.1111/jsr.13305

Arsintescu, L., Pradhan, S., Chachad, R. G., Gregory, K. B., Mulligan, J. B., & Flynn-Evans, E. E. (2022). Early starts and late finishes both reduce alertness and performance among short-haul airline pilots. *Journal of Sleep Research*, *31*(3), e13521. https://doi.org/10.1111/jsr.13521

Brezonakova, A. (2017). Pilot Burnout as a Human Factor Limitation. *Transportation Research Procedia*, *28*, 11–15. https://doi.org/10.1016/j.trpro.2017.12.163

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

Flaa, T. A., Bjorvatn, B., Pallesen, S., Zakariassen, E., Harris, A., Gatterbauer-Trischler, P., & Waage, S. (2022). Sleep and Sleepiness Measured by Diaries and Actigraphy among Norwegian and Austrian Helicopter Emergency Medical Service (HEMS) Pilots. *International Journal of Environmental Research and Public Health*, *19*(7), 4311. https://doi.org/10.3390/ijerph19074311

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., Signal, T. L., van den Berg, M. J., Mulrine, H. M., Jay, S. M., & Jim Mangie, C. (2013). 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

Gläsener, D., Post, J., Cyrol, D., & Sammito, S. (2023). Fatigue among Air crews on (Ultra)-Long-Range flights—A comparison of subjective fatigue with objective concentration ability. *Heliyon*, *9*(11), e21669. https://doi.org/10.1016/j.heliyon.2023.e21669

Goffeng, E. M., Wagstaff, A., Nordby, K.-C., Meland, A., Goffeng, L. O., Skare, Ø., Lilja, D., & Lie, J.-A. S. (2019). Risk of Fatigue Among Airline Crew During 4 Consecutive Days of Flight Duty. *Aerospace Medicine and Human Performance*, *90*(5), 466–474. https://doi.org/10.3357/AMHP.5236.2019

Goode, J. H. (2003). Are pilots at risk of accidents due to fatigue? *Journal of Safety Research*, *34*(3), 309–313. https://doi.org/10.1016/s0022-4375(03)00033-1

Lamond, N., Petrilli, R. M., 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

Li, J., Zhang, X., & Zhou, Y. (2023). The sustained attention characteristics of flight crews on exempt and non-exempt flights. *Chronobiology International*, *40*(7), 926–937. https://doi.org/10.1080/07420528.2023.2235011

Li, J., Zhou, Y., Zhang, X., & Fan, T. (2022). Fatigue during Long-Haul Flights of Different Crew Compositions under Exemption from Layover and Flight Time during COVID-19. *International Journal of Environmental Research and Public Health*, *19*(20), 13567. https://doi.org/10.3390/ijerph192013567

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

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

Powell, D. M. C., Spencer, M. B., Holland, D., Broadbent, E., & Petrie, K. J. (2007b). 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. (2010). Fatigue in airline pilots after an additional day's layover period. *Aviation, Space, and Environmental Medicine*, *81*(11), 1013–1017. https://doi.org/10.3357/asem.2804.2010

Roach, G. D., Petrilli, R. M. A., Dawson, D., & Lamond, N. (2012). Impact of layover length on sleep, subjective fatigue levels, and sustained attention of long-haul airline pilots. *Chronobiology International*, *29*(5), 580–586. https://doi.org/10.3109/07420528.2012.675222

Ruscitto, C., Ogden, J., & Ellis, J. G. (2023). To what extent is circadian phase predictive of subjective jet lag in long-haul cabin crew pre- and post-trip? *Applied Ergonomics*, *106*, 103882. https://doi.org/10.1016/j.apergo.2022.103882

Sallinen, M., Onninen, J., Ketola, K., Puttonen, S., Tuori, A., Virkkala, J., & Åkerstedt, T. (2021). Self-reported reasons for on-duty sleepiness among commercial airline pilots. *Chronobiology International*, *38*(9), 1308–1318. https://doi.org/10.1080/07420528.2021.1927071

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

Samel, A., & Wegmann, H. M. (1987). Desynchronization and internal dissociation in aircrew. *Ergonomics*. https://doi.org/10.1080/00140138708966032

Samel, A., Wegmann, H. M., & Vejvoda, M. (1997). Aircrew fatigue in long-haul operations. *Accident; Analysis and Prevention*, *29*(4), 439–452.

Sammito, S., Cyrol, D., & Post, J. (2022). Fatigue and Ability to Concentrate in Flight Attendants During Ultra-Long-Range Flights. *High Altitude Medicine & Biology*, *23*(2), 159–164. https://doi.org/10.1089/ham.2021.0173

Signal, T. L., Mulrine, H. M., van den Berg, M. J., Smith, A. A. T., Gander, P. H., & Serfontein, W. (2014). Mitigating and monitoring flight crew fatigue on a westward ultra-long-range flight. *Aviation, Space, and Environmental Medicine*, *85*(12), 1199–1208. https://doi.org/10.3357/ASEM.4034.2014

Sun, J., & Sun, R. (2022). Forecasting crew fatigue risk on international flights under different policies in China during the COVID-19 outbreak. *Frontiers in Public Health*, *10*, 996664. https://doi.org/10.3389/fpubh.2022.996664

Sun, J., Sun, R., Li, J., Wang, P., & Zhang, N. (2022). Flight crew fatigue risk assessment for international flights under the COVID-19 outbreak response exemption policy. *BMC Public Health*, *22*(1), 1843. https://doi.org/10.1186/s12889-022-14214-5

Sun, J.-Y., Liao, Y., Guo, H., & Jia, H. (2024). Fatigue risk assessment for flight crews flying across time zones in different directions to the east or west during the COVID-19 pandemic in China. *BMC Public Health*, *24*(1), 3561. https://doi.org/10.1186/s12889-024-20856-4

van den Berg, M. J., Signal, T. L., Mulrine, H. M., Smith, A. A. T., Gander, P. H., & Serfontein, W. (2015b). Monitoring and Managing Cabin Crew Sleep and Fatigue During an Ultra-Long Range Trip. *Aerospace Medicine and Human Performance*, *86*(8), 705–713. https://doi.org/10.3357/AMHP.4268.2015

Wegmann, H. M., Gundel, A., Naumann, M., Samel, A., Schwartz, E., & Vejvoda, M. (1986). Sleep, sleepiness, and circadian rhythmicity in aircrews operating on transatlantic routes. *Aviation, Space, and Environmental Medicine*, *57*(12 Pt 2), B53-64.

Wu, L. J. (2013). *Evidence Based Fatigue Risk Management during 24/7 Operations: Objective Assessment of Pilots' Sleep, Performance, and Fatigue during Ultra Long Range and Long Range Flights.* Washington State University.

Yang, S. X., Cheng, S., Sun, Y., Tang, X., & Huang, Z. (2024). Circadian Disruption in Civilian Airline Pilots. *Aerospace Medicine and Human Performance*, *95*(7), 381–389. https://doi.org/10.3357/AMHP.6316.2024

Link to BAM technical fact sheet: BAM_Tech_Fact_Sheet_2.3.pdf
Link to SAFTE implementation code: GitHub - InstituteBehaviorResources/SAFTEr: R package for researchers to apply the SAFTE model to sleep data for comparison to other collected observations.

European Union Aviation Safety Agency

Konrad-Adenauer-Ufer 3

50668 Cologne

Germany

Mail EASA.research@easa.europa.eu

Web www.easa.europa.eu