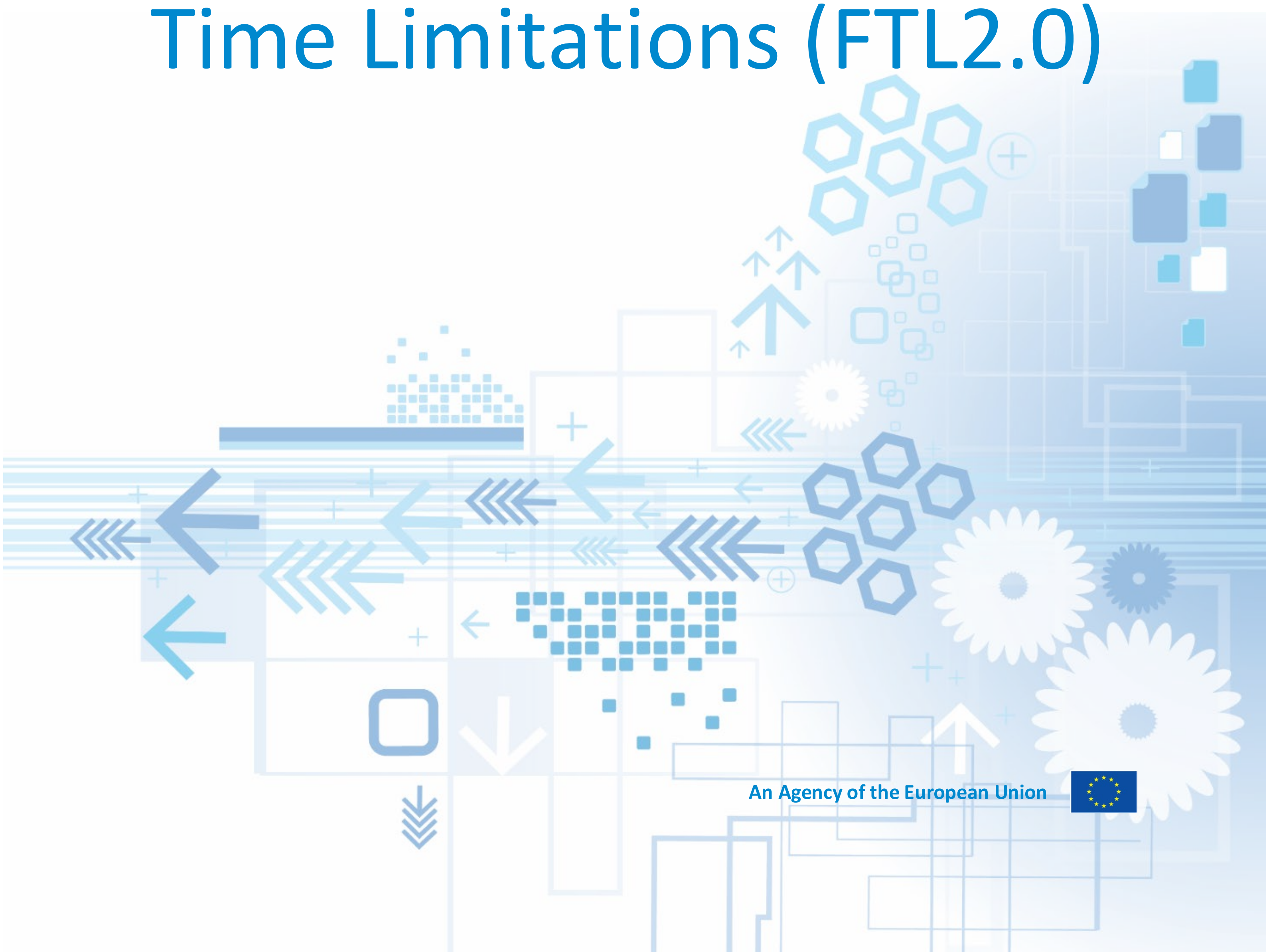
EASA
European Union Aviation Safety Agency

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

RESEARCH PROJECT [EASA.2019.C31]

DELIVERABLE 2.5: MERGE OF THE RESULTS OF THE CURRENT AND PREVIOUS CONTRACT

Effectiveness of Flight Time Limitations (FTL2.0)



An Agency of the European Union

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Union Aviation Safety Agency (EASA). Neither the European Union nor EASA can be held responsible for them.

This deliverable has been carried out for EASA by an external organisation and expresses the opinion of the organisation undertaking this deliverable. It is provided for information purposes. Consequently, it should not be relied upon as a statement, as any form of warranty, representation, undertaking, contractual, or other commitment binding in law upon the EASA.

Ownership of all copyright and other intellectual property rights in this material including any documentation, data and technical information, remains vested to the European Union Aviation Safety Agency. All logo, copyrights, trademarks, and registered trademarks that may be contained within are the property of their respective owners. For any use or reproduction of photos or other material that is not under the copyright of EASA, permission must be sought directly from the copyright holders.

No part of this deliverable may be reproduced and/or disclosed, in any form or by any means without the prior written permission of the owner. Should the owner agree as mentioned, then reproduction of this deliverable, in whole or in part, is permitted under the condition that the full body of this Disclaimer remains clearly and visibly affixed at all times with such reproduced part.

DELIVERABLE NUMBER AND TITLE: D2.5. Merge of the results of the current and previous contract (DRAFT)
CONTRACT NUMBER: EASA.2019.C31
CONTRACTOR / AUTHOR: NLR / Alwin van Drongelen
IPR OWNER: European Union Aviation Safety Agency
DISTRIBUTION: Public

APPROVED BY:	AUTHORS	REVIEWER	MANAGING DEPARTMENT
	M. Sallinen (FIOH) T. Åkerstedt (SU) D. Fischer (DLR) D. Aeschbach (DLR) A. Roelen (NLR) L. Marsman (NLR) A. van Drongelen (NLR)	A. Maij (NLR)	A. Rutten (NLR)
		EASA: Irina Petrova, EASA Flight Standards Directorate Scientific Committee: Barbara Stone, Alexandra Holmes, Kristjof Tritschler	

DATE: 10 April 2025

SUMMARY

Problem area

The research study FTL 2.0 aims to perform a review of the effectiveness of current flight and duty time limitations and rest requirements contained in Commission Regulation (EU) No 965/2012 in maintaining acceptable levels of aircrew alertness. 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

The objective of Task 2.5 is to bring together the results of the previous phase (FTL1) and the results of the current contract (FTL2.0). These results are discussed in the context of safety performance metrics that can be used to assess the effectiveness of the current flight and duty time limitations of Annex III of Commission Regulation (EU) No 965/2012 and the effectiveness of controlled rest in maintaining acceptable levels of aircrew alertness. The outcomes of this deliverable are:

- A synopsis of the results of FTL1 and FTL2.0;
- A list of safety performance metrics chosen to assess how effective the current flight and duty time limitations of Commission Regulation (EU) No 965/2012 are; and
- A description of safety performance metrics chosen to assess the effectiveness of controlled rest in improving aircrew alertness.

Results and Application

The results of FTL1 showed that all disruptive schedules examined were associated with a higher probability (relative frequency) of high fatigue (Karolinska Sleepiness Scale (KSS) ≥7) at the last ToD (last sector of the FDP) (range 0.22 – 0.42) as compared to daytime FDPs (0.16), with the most prominent increase occurring in the context of late finishing (0.31) and night (night: 0.32; night >10 hours: 0.42) FDPs. Also, the subcategory of night FDPs with an end time of 06:00 or later showed a high probability of KSS ≥7 at the last ToD (0.42).

As regards to the FTL2.0 study, it showed that FDPs longer than 13 hours (FDPs >13 hours) starting at the most favourable time of the day (start time between 06:00h and 13:29h) are associated with a high probability (relative frequency) of high fatigue at the last ToD (0.40). No evidence was found that this result could have been attributable to disruptive schedules, such as late finishes or night FDPs.

In general, a longer FDP duration (FTL1 and FTL2.0) and a higher number of sectors (FTL1 and FTL2.0) were associated with a higher probability of high fatigue at last ToD.

When examining the 18 hours awake time cap in the context of other-than-airport standby, it was found that self-rated fatigue during FDPs assigned during these standby periods was associated more closely with the amount of prior sleep and the sleep-to-awake time ratio than the time spent awake per se. However, these results were largely based on non-disruptive assignments only.

The study on controlled rest (CR) found that CR is a frequently used fatigue mitigation strategy in pilots. In the current dataset it was shown that CR was particularly used during overnight long-haul flights with multiple time zone crossings. Scientific literature showed that CR can be successful in mitigating unexpected onset of fatigue in the cockpit. This was confirmed by the data of the current study.

The following conclusions about safety performance metrics can be drawn from the results of FTL1 and FTL2.0 in the context of assessing the effectiveness of current flight and duty time limitations in maintaining acceptable levels of aircrew alertness[1]:

- The Karolinska Sleepiness Scale (KSS) can be used as a safety performance metric to assess roster-induced fatigue. Especially the probability (relative frequency) of KSS ≥7 and KSS ≥8 ratings is useful for this purpose. It is recommended to use both thresholds. KSS ≥7 may include ratings indicating increased fatigue without severe self-reported difficulties staying awake (KSS =7). In contrast, KSS ≥8 includes only ratings indicating increased fatigue with such self-reported difficulties[2].
- The Samn-Perelli Fatigue scale (SP) can be used as a safety performance metric to assess roster-induced fatigue. Especially the probability (relative frequency) of SP ≥5 and SP ≥6 ratings is useful for this purpose. It is recommended to use both thresholds. SP ≥5 may include ratings indicating increased fatigue without severe self-reported difficulties concentrating or functioning (SP =5). In contrast, SP ≥6 includes only ratings indicating increased fatigue with such self-reported difficulties.
- Psychomotor Vigilance Task (PVT) outcome measures (reaction time, response speed) show to be insensitive to changes in FDP characteristics, such as start time, end time, duration, a number of sectors, and therefore cannot be considered as useful safety performance metrics for assessing the effectiveness of current FTL in maintaining acceptable levels of aircrew alertness, in unsupervised conditions and in real-world aviation settings.
- The amount of sleep in the preceding 24h, and time awake appear to be useful outcome measures for indirectly assessing roster-induced fatigue. Shorter sleep and longer awake time are associated with a higher probability of high self-rated fatigue.
- Controlled rest may not completely eliminate high fatigue, but could be used to reduce the likelihood of high fatigue during critical phases of specific types of flights (long-haul, inbound overnight duties involving the crossing of multiple time zones). This result was found using KSS ≥7 as fatigue outcome measure.

---

[1] Note that all conclusions on the effectiveness of the current flight and duty time limitations as such, and recommendations for their further development are presented in the FTL2.0 deliverables D2.6 and D2.7.

[2] Note that in terms of performance, almost half of the individuals in a driving experiment show a significantly increased risk of an accident or erratic driving a few minutes after rating 7 on the KSS, whereas almost all of them do so after rating 8 or 9 (Anund et al 2008; Åkerstedt et al 2013). Therefore, also KSS =7 can be considered a risk indicator, at least with reference to the driving task.

CONTENTS

ABBREVIATIONS

ACRONYM	DESCRIPTION
Bon	Blocks on
CI	Confidence Interval
CR	Controlled Rest
D	Deliverable
DLR	German Aerospace Centre
EASA	European Union Aviation Safety Agency
EU	European Union
FDP	Flight Duty Period
FTL	Flight and duty time Limitations and rest requirements
h	hour
KSS	Karolinska Sleepiness Scale
NLR	Royal NLR - Netherlands Aerospace Centre
PVT	Psychomotor Vigilance Task
RT	Reaction Time
SP	Samn Perelli
ToD	Top of Descent
WOCL	Window of Circadian Low

1. Introduction

1.1 Background

During the adoption in 2014 of the next generation of EU's Flight Time Limitation (FTL) and rest requirements for scheduled and charter airline operations, the European Parliament and the Commission instructed EASA to perform a continuous review of the effectiveness of those requirements.

Article 9b of Commission Regulation (EU) No 965/2012 stipulates that such review shall include an assessment of the impact on aircrew alertness of the following aircrew duty periods:

1. Duties of more than 13 hours at the most favourable time of the day;
2. Duties of more than 10 hours at the less favourable time of the day;
3. Duties of more than 11 hours for crew members in an unknown state of acclimatisation;
4. Duties including a high level of sectors (more than 6);
5. On-call duties such as standby or reserve followed by flight duties; and
6. Disruptive schedules.

1.2 Project description

The first phase of the study on Effectiveness of Flight Time Limitation (MOVE/C2/2016-360), (hereinafter referred to as FTL1), assessed the impact of Flight Duty Periods (FDPs) of more than 10 hours at the less favourable time of the day and disruptive schedules (i.e., early starts, late finishes, and night duties) on aircrew alertness.

The objective of the second phase of the study on Effectiveness of Flight Time Limitation (EASA.2019.C31), (hereinafter referred to 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.
- Controlled rest (CR): this includes an analysis of the conditions and circumstances under which aircrew members take CR.

Two series of deliverables will be provided in FTL2.0:

- 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 Task 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 2.6 and 2.7 on the analysis of effectiveness of prescriptive FTL and the conclusions and recommendations.

1.3 This deliverable

The objective of Task 2.5 was to combine the results of the previous phase of the study (FTL1) and the results of the current contract (FTL2.0) with a view to establish generally applicable safety performance metrics to assess the effectiveness of the current FTL of Regulation (EU) 965/2012 and of controlled rest in maintaining acceptable levels of aircrew alertness. This deliverable D2.5 therefore includes:

- A synopsis of the results of the previous contract and the current work;
- A list of safety performance metrics chosen to assess how effective the current flight and duty time limitations of Commission Regulation (EU) No 965/2012 are; and
- A description of safety performance metrics chosen to assess the effectiveness of controlled rest in improving aircrew alertness.

1.4 Approach

The results of FTL1 and FTL2.0 are summarised and put together[1] using the same or comparable fatigue metrics from these two projects. The results of CR obtained in FTL2.0 are summarised in a separate subsection, since these were not specifically investigated in FTL1.

The primary metric used is the nine-point Karolinska Sleepiness Scale (KSS) and the primary outcome measure the probability (relative frequency) of high fatigue, defined as a rating ≥7 given at the last ToD of the FDP (Åkerstedt et al., 2014; Åkerstedt & Gillberg, 1990; Ingre at al., 2006). Given the objective of this deliverable (see section 1.2), this metric and outcome measure can be considered highly relevant. The relationship between the KSS and Samn-Perelli Fatigue Scale (SP) scale (Powell et al., 2011; Samn & Perelli, 1982) is addressed in a separate subsection. The results of the Psychomotor Vigilance Task (PVT) (Arthurs et al., 2021; Basner et al., 2011; Benderoth et al., 2021; Elmenhorst et al., 2012) obtained in study FTL1 and study FTL2.0 are also summarised and put together.

The results of the FTL1 and FLT2.0 studies are also summarised and put together to provide a comprehensive overview of the main predictors of high fatigue at last ToD. These results can be used to identify which metrics, in addition to direct fatigue metrics, have the largest potential for assessing the effectiveness of the current flight and duty time limitations.

Overall, our approach enables to assess the suitability of fatigue metrics to evaluate the effectiveness of the current FTL of Regulation (EU) 965/2012 and of controlled rest in maintaining acceptable levels of aircrew alertness. These topics are also addressed in the discussion section (Chapter 4) in light of the main results of FTL1 and FTL2.0.

---

[1] **FDPs of interest in FTL1**

- FDPs of more than 10 hours at the less favourable time of day
    - FDPs > 10 hours that encroach any portion of the period between 02:00h and 05:00h.
- Disruptive schedules
    - Early starts, late finishes, nights, and combinations thereof.

FDPs of interest in FTL2.0

- FDPs of more than 13 hours at the most favourable time of the day
    - FDPs that start between 06:00h and 13:29h, include 1-2 sectors, and are flown in a known state of acclimatisation
- FDPs of more than 11 hours for crew members in an unknown state of acclimatisation
- FDPs including a high level of sectors (>6)
- FDPs assigned during other than airport standby

In addition to these four FDPs of interest, CR was examined in FTL2.0.

NB. All FDPs of interest, other than CR, were studied in both cockpit and cabin crew. For flight crew, only non-augmented crews were considered.

2. Methods

2.1 Main methods and metrics used in FTL1 and FTL2.0

The main quantitative metrics measured in the field campaigns of study FTL1 and study FTL2.0 were: on-duty fatigue, sleep, and awake time. Self-rated on-duty fatigue was measured using the KSS as well as the SP (see below for details on the scales). The primary outcome measure of the KSS was the probability of a high fatigue rating (threshold ≥7) at the last ToD of the FDP and a secondary outcome measure the probability of a very high fatigue rating (threshold ≥8) at the same flight phase. The corresponding outcome measures of the SP were the probability of a high fatigue rating (threshold ≥5) and that of a very high fatigue rating (threshold ≥6) at the last ToD of the FDP.

Karolinska Sleepiness 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 to keep awake, fighting sleep

Samn-Perelli 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.

The term 'probability' can have two meanings in this deliverable. When it is used in descriptive statistics it means relative frequency, that is, the ratio of the number of cases (e.g. the number of KSS ratings ≥7) to the total number of observations (e.g. all KSS ratings). When the term is used in the context of a statistical regression analysis, it means the predicted probability is based on the estimates provided by the regression model. In both cases, the value of probability falls between 0 and 1.

Vigilance was measured by two versions of the PVT. In study FTL1, a 5-min version was being performed approximately 10 minutes before the last ToD of the FDP. The primary outcome was mean reaction time (ms). To better align with the goals of study, in FTL2.0 a 3-min version of the PVT was being performed soon after the last sector of the FDP. The primary PVT outcome measure was mean response speed (mean of 1/reaction time). In both studies, also the number of lapses was being measured with reaction times longer than 500ms (FTL1) and longer than 355ms (FTL2.0). These thresholds were established on the basis of previous scientific publications (Grant et al. 2017; Antler et al. 2022).

Sleep logs were being used to collect subjective data on sleep/wake cycles and sleep quantity and quality in both studies. The sleep log included items on bed time, lights off time, wake-up time, sleep quality, feelings of restedness, and sleep location. The primary outcome measures were hours of sleep in the past 24 hours and hours of being awake at the end of the FDP.

The outcome measures used for describing and classifying FDPs were start time, end time, duration, and the number of sectors. All FDPs included in the analyses were operated by a non-augmented flight crew. All data were collected by the same subjective log used to collect the sleep-wake data.

All the methods used in FTL1 and FTL2.0 are more specifically presented in the deliverables D2.2 (Definition of the Data Collection Process - FTL1) and D2.1 (Data collection Scope and Process - FTL2.0).

3. Results

3.1 Synopsis of the main results of the FDPs of interest

3.1.1 Probability of high KSS rating at last ToD

This subsection focuses on the results of the probability of high (KSS ≥7) and very high (≥8) ratings given at last ToD. The former threshold was used as the primary fatigue outcome measure and the latter as a secondary one in both the FTL1 and FTL2.0.

Figures 3-1 and 3-2 show the probability (relative frequency) of high fatigue (KSS ≥7) or very high fatigue (KSS ≥ 8) at last ToD for each FDP of interest respectively. These results suggest that FDPs which either start or end outside the daytime window (07:00h – 22:59h) are associated with an elevated probability for high fatigue at last ToD compared with FDPs starting and ending inside that window. In addition, the results suggest that the same holds for FDPs with a high number of sectors and FDPs of more than 11 hours where crew members are in an unknown state of acclimatisation, although the confidence interval (CI) of the latter FDPs is wide and overlaps that of the daytime FDPs. However, it is important to note that in the categories of FDP >13h starting at the most favourable time of the day, FDP with >6 sectors, and FDP >11h for crew members in an unknown state of acclimatisation, the majority of FDPs occurred at least partly outside the 07:00 – 20:00 window. Moreover, any direct comparison between the FDP categories should be made with caution because there are differences in the number of observations per individual crew member and there is a limited number of observations within some FDP types (e.g. 6 observations for FDP >11h for crew members in an unknown state of acclimatisation, and 33 observations for FDP >13h starting at the most favourable time of the day).

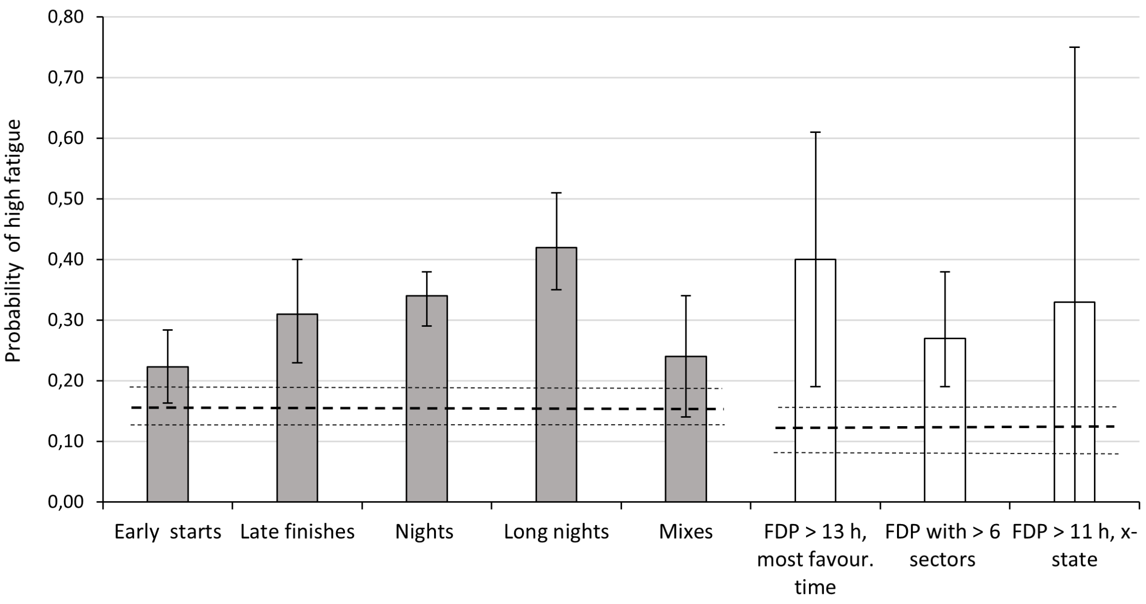


Figure 3-1. *Probability (95% CI) of high fatigue (KSS ≥7) at last ToD within the FDPs of interest. The grey bars denote the FDPs of interest in FTL1 and the white bars the FDPs of interest in FTL2.0. The vertical lines indicate 95% CIs of the FDPs of interest. The thick dashed horizontal lines denote the corresponding probability within all daytime FDPs collected in the respective FTL study and the thin dashed lines - the related 95% CI. The number of observations by FDP type are as follows: early starts: 197, late finishes: 156, nights: 494, consecutive nights: 92, long nights (>10h): 146, mixed combinations of disruptive schedules: 79, FDP >13h starting at the most favourable time of the day: 25, FDP with >6 sectors: 94, and FDP >11h for crew members in an unknown state of acclimatisation: 6.*

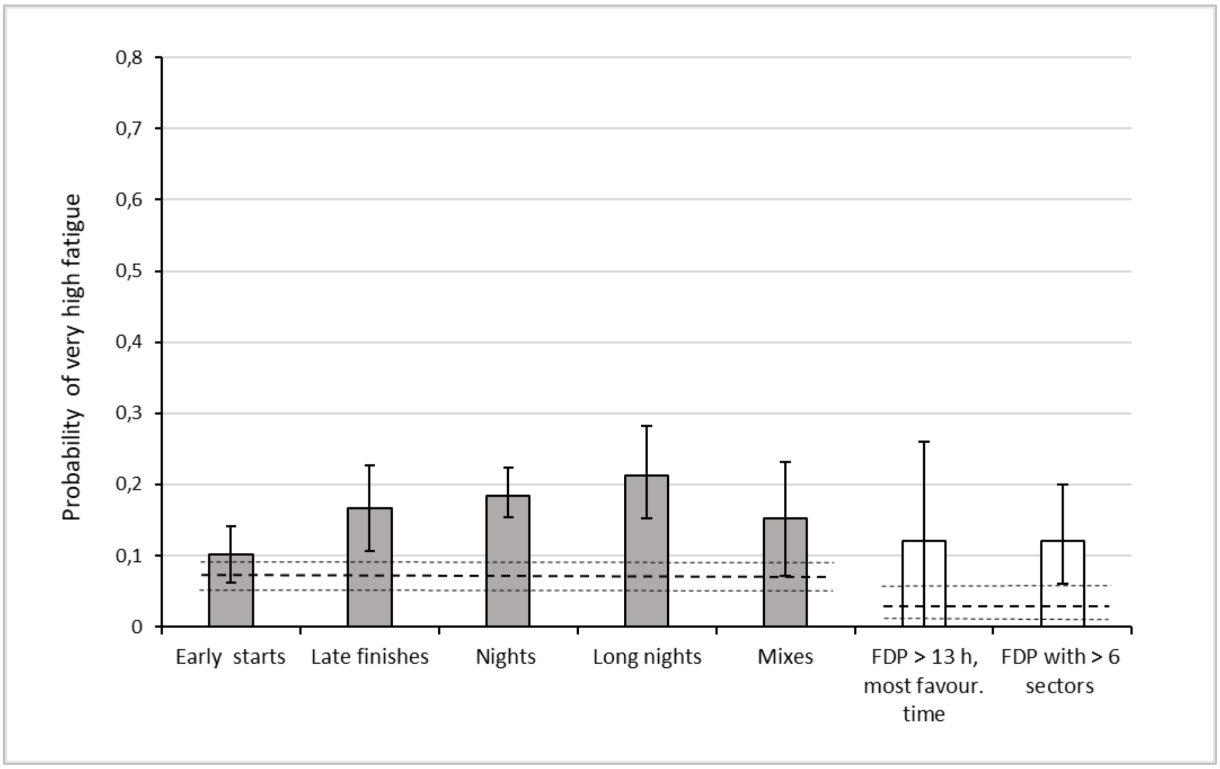


Figure 3-2. *Probability (95% CI) of very high fatigue (KSS ≥8) at last ToD within the FDPs of interest. The grey bars denote the FDPs of interest in FTL1 and the white bars the FDPs of interest in FTL2.0. The vertical lines indicate 95% CIs of the FDPs of interest. The thick dashed horizontal lines denote the corresponding probability within all daytime FDPs collected in the respective FTL study and the thin dashed lines the related 95% CI. The number of observations by FDP type are as follows: early starts: 197, late finishes: 156, nights: 494, consecutive nights: 92, long nights (>10h): 146, mixed combinations of disruptive schedules: 79, FDP >13h flown at the most favourable time of the day: 25, FDP with >6 sectors: 94, and FDP >11h for crew members in an unknown state of acclimatisation: 6. There were no KSS ratings of 8 or 9 for FDPs >11h for crew members in an unknown state of acclimatisation (hence, no bar is shown).*

Figures 3-3 and 3-4 show the probability of high fatigue (KSS ≥7 and KSS ≥8) at last ToD for those FDPs of interest which were compared to control FDPs which were other-than-daytime FDPs.

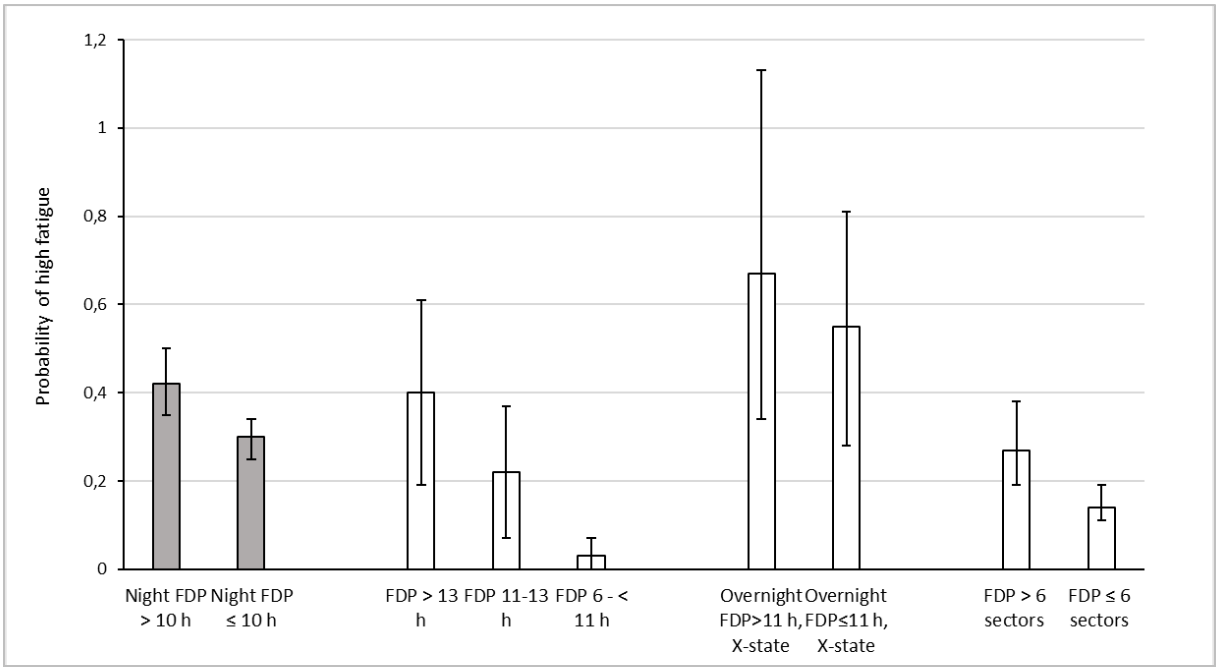


Figure 3-3. *Probability (95% CI) of high fatigue (KSS ≥7) at last ToD for (a) night FDP >10h vs night FDP ≤10h, (b) FDPs of more than 13 hours vs FDPs 11-13 hours vs FDPs 6 - <11 hours starting at the most favourable time of the day, c) overnight FDPs of > 11 hours vs ≤11 for crew members in an unknown state of acclimatisation, d) FDPs including >6 sectors vs FDPs including ≤ 6 sectors. In connection with FDPs flown in an unknown state of acclimatisation, "overnight" does not refer to the biological night of the crew member but to local night hours, assessed by an FDP end time between or FDP coverage of 2:00-5:59 at the arrival time zone, which was always home base in the present dataset. The grey bars denote the FDPs collected in FTL1 and the white bars those collected in FTL2.0.*

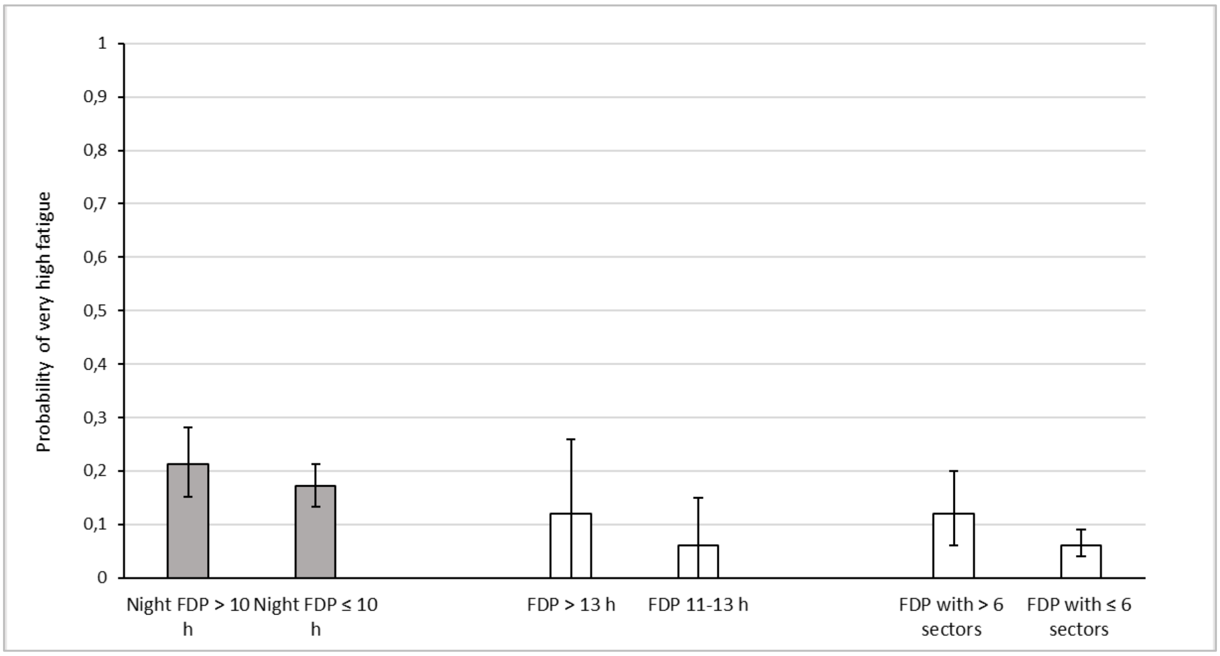


Figure 3-4. *Probability (95% CI) of very high fatigue (KSS ≥8) at last ToD for (a) night FDP >10h vs night FDP ≤10h, (b) FDPs of more than 13 hours vs FDPs 11-13 hours vs FDPs 6 - <11 hours starting at the most favourable time of the day, c) FDPs including >6 sectors vs FDPs including ≤6 sectors. The grey bars denote the FDPs collected in FTL1 and the white bars those collected in FTL2.0. There were no KSS ratings of 8 or 9 for FDPs 6 - <9h (control FDP for FDP >13h) or FDPs >11h for crew members in an unknown state of acclimatisation.*

Overall, the results shown in Figures 3-1 to 3-4 suggest that the probability of high and very high fatigue responds to variations in FDP characteristics such as duration, timing (start and end time), and the number of sectors.

The statistical comparisons conducted in FTL1 showed that the predicted probability of high fatigue (KSS ≥7) at last ToD

a) did not differ significantly between night FDPs >10 hours and night FDPs ≤10 hours;
b) was significantly higher for late finish and night FDPs but less evidently so for early start FDPs in comparison with non-disruptive schedules (i.e., daytime FDPs)

The statistical comparisons (based on multivariable analyses *unless* otherwise stated) conducted in FTL2.0 showed that the predicted probability of high fatigue (KSS ≥7) at last ToD[2]:

a) was higher for FDPs >13 hours and FDPs 11-13 hours than FDPs 6 - <11 hours (in the single variable analysis only) but no difference was found between FDPs >13 hours and FDPs 11-13 hours (FDP start time between 06:00h and 13:29h in all cases) (as shown in Figure 3-5).
b) was higher for FDPs ≥9 hours than for FDPs <9 hours flown overnight in an unknown state of acclimatisation (as shown in Figure 3-6).
c) was higher for FDPs including >6 sectors than FDPs including ≤6 sectors ((as shown in Figure 3-7).

Please note that the results presented in Figures 3-5 to 3-7 are based on estimates of the multivariable analyses and adjusted for a number of factors for which they do not match completely with the descriptive results shown in Figures 3-1 to 3-4. The statistical results are presented in more detail in the FTL1 and FTL2.0 Deliverables 2.3.

Overall, these statistical results show that the probability of high fatigue (KSS ≥7) is a function of FDP characteristics such as duration, timing (start and end time), and the number of sectors.

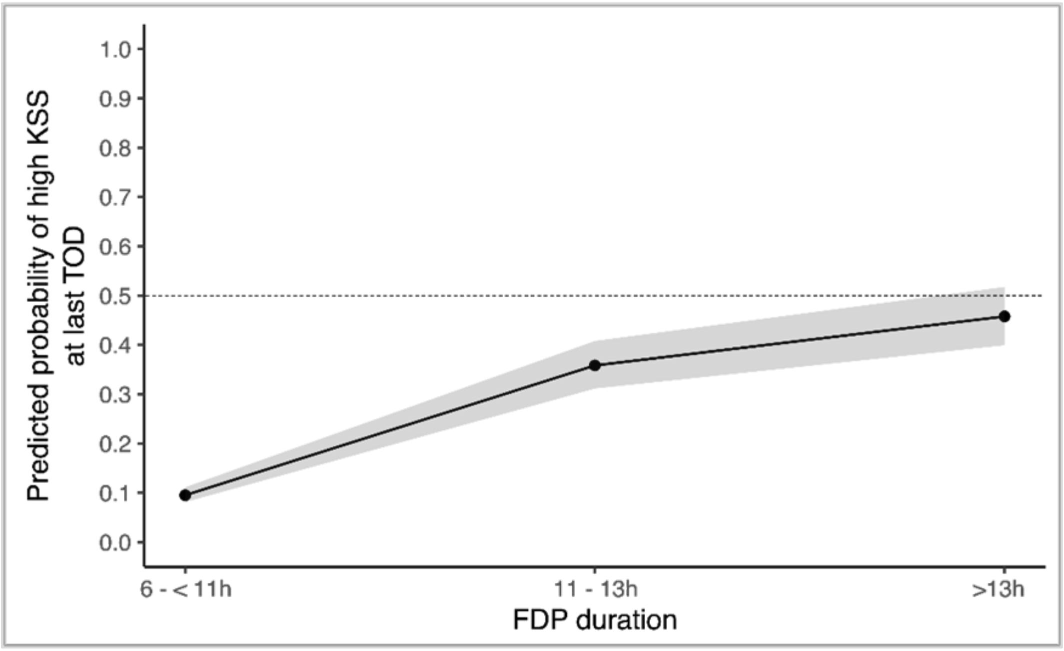


***Figure 3-5**. Predicted probability of high fatigue (KSS ≥7) at last ToD by FDP duration. The grey area indicates the predicted 95% CI. The data concerns 148 FDPs and 86 subjects. All FDPs start between 06:00h and 13:29h, include 1-2 sectors, and are flown by non-augmented flight crew who is in a known state of acclimatisation. Adjusted to the sub-sample mean values for the continuous variables; female, cabin crew, one sector, and non-disruptive schedule for the categorical variables.*

---

[2] For multi-sector FDPs, KSS ratings were given shortly after the end of the last sector of the FDP instead of last ToD.

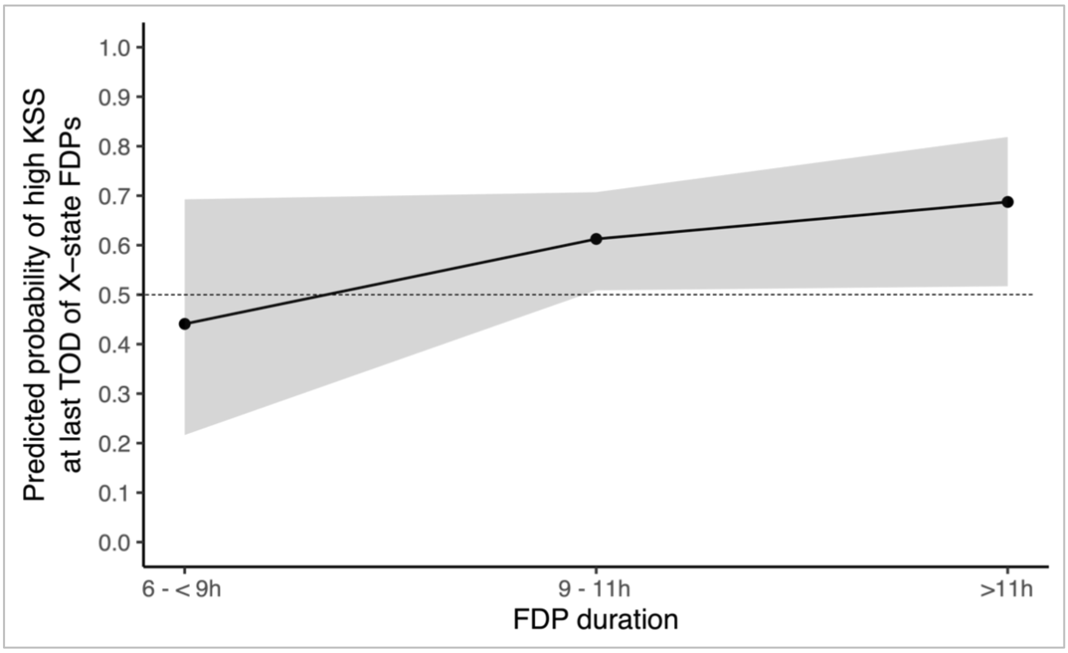


Figure 3-6. *Predicted probability of high fatigue (KSS ≥7) at last ToD of X-state FDPs. The grey area indicates the predicted 95% CI. Adjusted to sample mean values for continuous variables; female; cabin crew; inbound flight of a westward rotation; overnight-FDP; one sector.*

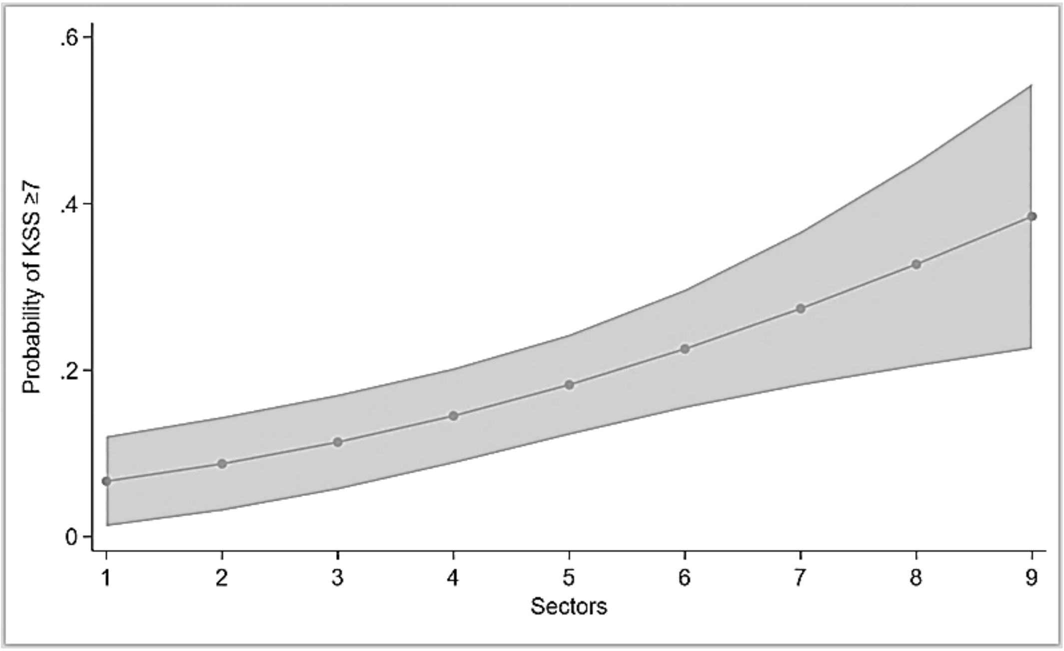


Figure 3-7. *Probability of KSS ≥7 at last sector as a function of the number of sectors. Mean ±95 CI (shaded).N=77, observations/participant =4.7. The number of individuals and total number of observations (N/obs) per number of sectors: 1 sector=27/31, 2=9/11, 3=18/27, 4=49/94, 5=43/72, 6=36/55, 7=21/54, 8=20/28, 9=3/3. Adjusted to sample mean values for continuous variables and to the value of the reference group for categorical variables.*

3.1.1.1 Categorization of night FDPs

In FTL1, night FDPs were divided in three sub-categories to see how these sub-categories should be addressed in fatigue mitigation. Figure 3-8 shows the probability of high fatigue (KSS ≥7) at last ToD for early starts (start time between 05:00 and 06:59), late finishes (end time between 23:00 and 01:59), night type (nights) 1 (start time between 02:00 and 04:59), night type (nights) 2 (start time 01:59 or earlier, end time between 02:00 and 04:59) and night type (nights) 3 (start time 01:59 or earlier, end time 06:00 or later) with daytime FDPs (start time 7:00 or later and end time 22:59 or earlier) as reference. Each FDP showed a larger probability of high fatigue, particularly for late finish, night, and night type 3 (FDPs with start time 01:59 or earlier, end time 06:00 or later).

The statistical analysis conducted showed that the probability of high fatigue at last ToD was increased in each FDP category as compared to non-disruptive FDPs with the most pronounced increase in the FDP category of night 3 (start time 01:59 or earlier, end time 06:00 or later) (Odds Ratio = 8.04). The statistical results based on single- and multivariable regression analyses are presented in more detail in FTL1 deliverable D2.3.

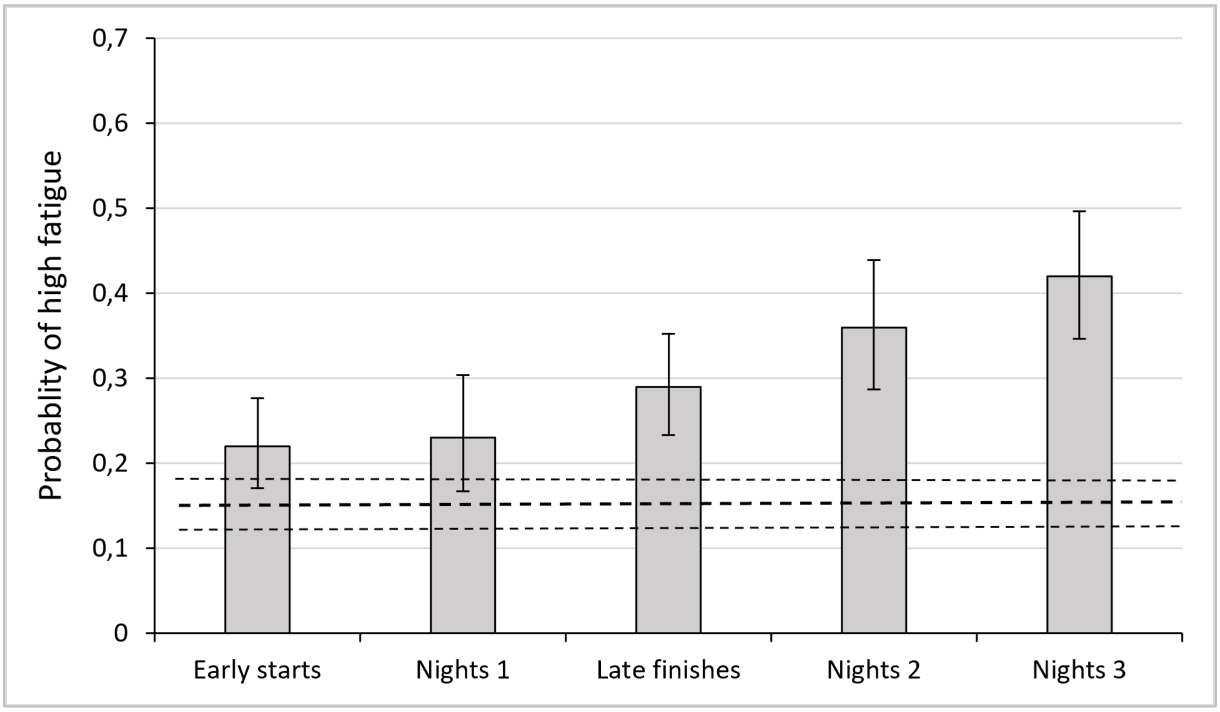


Figure 3-8. *Probability of high fatigue (KSS ≥ 7) at ToD for early starts (start time 05:00 - 06:59), nights 1 (start time 02:00 - 04:59), late finishes (end time 23:00 - 01:59), nights 2 (start time 01:59 or earlier, end time 02:00 - 04:59) and nights 3 (start time 01:59 or earlier, end time 06:00 or later). The thick dashed horizontal line denotes the corresponding probability of all daytime (whole FDP between 7:00 and 22:59) FDPs collected in FTL1 and the thin dashed lines the related 95% CI.*

3.1.1.2 Breakdown of FDPs >13 hours starting at the most favourable time of the day by FDP type

To get a better understanding of FDPs of more than 13 hours starting at the most favourable time of the day, an attempt was made to break these FDPs down by their ending times. The reason for this supplementary analysis, not published in the previous deliverables, was that these FDPs included not only FDPs ending at 22:59 or earlier, but also FDPs ending between 23:00 and 01:59 and those falling into the night FDP category. An attempt was made here to analyse descriptively the probability of high fatigue at last ToD of FDPs >13 hours separately for these FDP types. It was expected that the probability of high and very high fatigue would remain lower for FDPs ending no later than 22:59 than for FDPs ending later.

Figures 3-9a (top) and 3-9b (bottom) show the probability of high fatigue (KSS ≥7) and very high fatigue (KSS ≥8) at last ToD for FDPs >13 hours broken down by ending times. These descriptive results show that the relatively high probability of high fatigue at last ToD for FDPs >13 hours “starting at the most favourable time of the day” (see Figures 3-1 and 3-2) is not attributable to those FDPs that end very late or even encroach the window of 02:00 – 04:59 in the present data. However, it is premature to draw any conclusion on the role of the ending time of the FDP (late finish, night, or not disruptive FDPs) in the probability of high fatigue observed due to the low number of observations (see the legend of Figures 3-9a and 3-9b (e.g., 6 observations of daytime FDPs >13 hours).

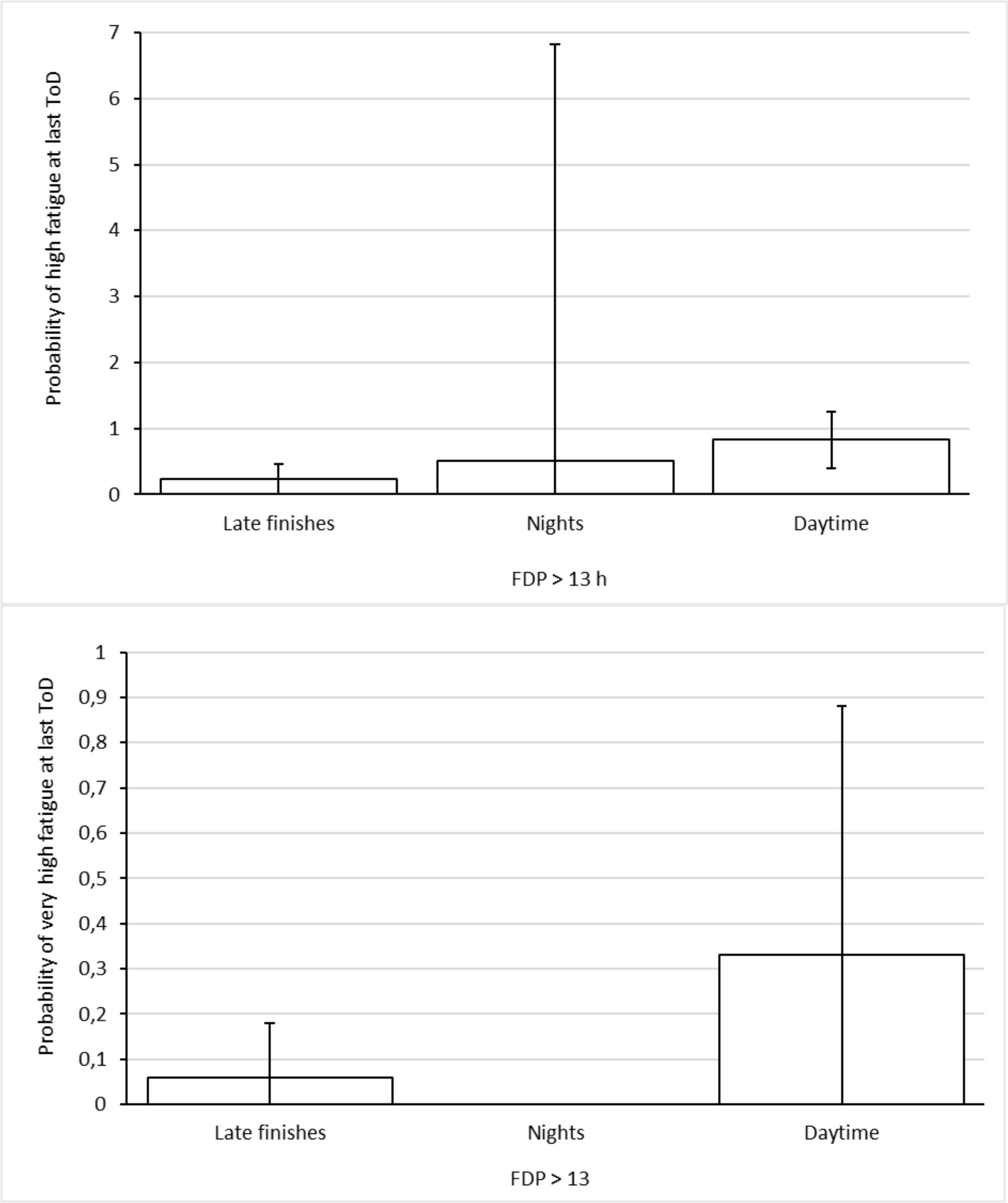


***Figure 3-9a (top) and 3-9b (bottom)**. Probability (95% CI) of high fatigue (KSS ≥7) (a) and very high fatigue (KSS ≥8) (b) at last ToD for FDPs >13 hours broken down by FDP type (late finish - end time 23:00h – 01:59h (early type country) or 00:00h – 01:59h (late type country); night (any portion of the FDP between 02:00h – 04:59h); daytime (whole FDP between 06:00h – 22:59h (early type country) or FDP between 07:00h – 23:59h (late type country)). The number of observations by FDP duration and FDP type are as follows: a) late finishes 17, nights 2, daytime 6. Note the difference in the scale of the y-axis between the figures a and b.*

3.1.1.3 FDPs assigned during other than airport standby

One of the duties of interest in FTL2.0 were FDPs that were not assigned in advance in duty rosters, but during other-than-airport standby periods. Given the unpredictable nature of these FDPs, there are variations within these FDPs in terms of duration, timing and the number of sectors. The main aim of this part of the study was to examine the effectiveness of an 18-hour cap on awake time. This cap on the maximum amount of awake time applies to the combined duration of the other-than-airport standby period and the assigned FDP.

Figure 3-10 shows the association between the time spent awake at the end of each assigned FDP and the KSS rating given at last ToD of these FDPs in the dataset. The association did not reach statistical significance and only one in sixteen KSS ratings indicated high fatigue (KSS =7). However, there was a significant association between the amount of sleep in the past 24 hours and KSS ratings given at last ToD of these FDPs (Figure 3-11), and the same applied to the sleep-to-awake time ratio and the corresponding KSS ratings.

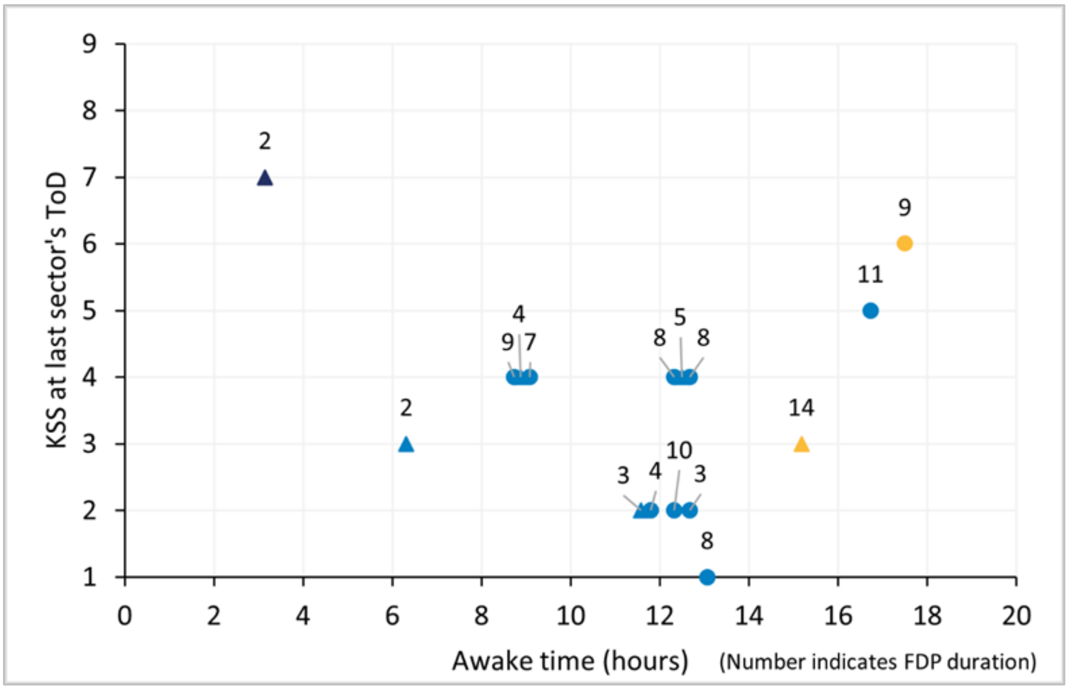


Figure 3-10. *Scatter plot between prior time awake and the KSS ratings given at the last ToD of the assigned FDPs. Number of participants is 12 (8 pilots, 4 cabin crew) and number of observations (FDPs) is 16. The numbers above each data point denote FDP duration in full hours. Legends: early starts (dark blue), non-disruptive (blue), late finish (yellow); ▲ 1 sector FDP, ● 2 sector FDP.*

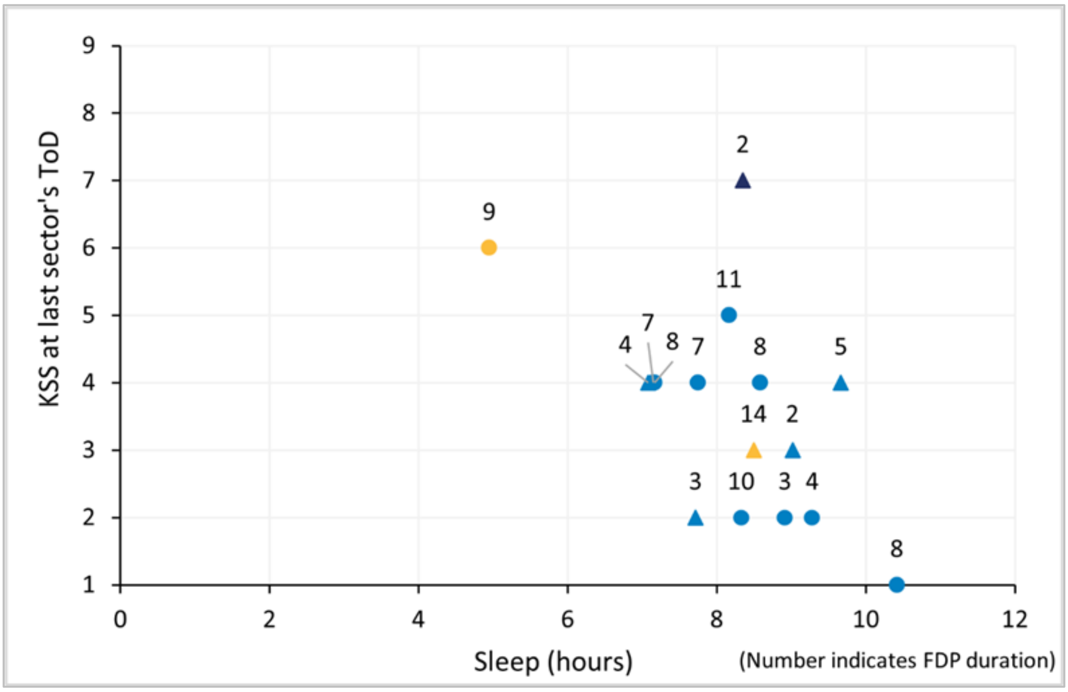


Figure 3-11. *Scatter plot between the amount of prior sleep and the KSS ratings given at the last ToD of the assigned FDPs. Number of participants is 12 (8 pilots, 4 cabin crew) and number of observations (FDPs) is 16. The numbers above each data point denote FDP duration in full hours. Legends: early start (dark blue), non-disruptive (blue), late finish (yellow); ▲ 1 sector FDP, ● 2 sector FDP.*

3.1.2 Karolinska Sleepiness Scale vs Samn-Perelli Fatigue Scale

As mentioned above, the KSS Sleepiness Scale was used along with the SP Fatigue Scale in both FTL1 and FTL2.0 studies. This subsection addresses the relationship between these two scales, as derived from the collected KSS and SP data.

In both FTL1 and FTL2.0, the (Pearson) correlation between KSS and SP ratings given at ToD was found to be high. In FTL1, there was a strong positive correlation between KSS and SP ratings (*r* = .87) (Figure 3-12). In FTL2.0, the corresponding correlation was the same (*r* = .87, p< .001) when using the multi-sector dataset for the analysis. In addition, having KSS ratings as the predictor and SP ratings as the response variable, a regression coefficient of 0.59 (se ± 0.021, intercept 0.98 ± 0.032) was found (Figure 3-13). This means that SP ratings increased on average by 0.59 units for each one-unit increase in KSS (intercept 0.59). Thus, KSS =7 translates to 5.1 and KSS =9 to 6.2 on the SP scale. Note that the shaded standard error marker in Figure 3-13 is very 'narrow'. This indicates the close relationship between the two scales.

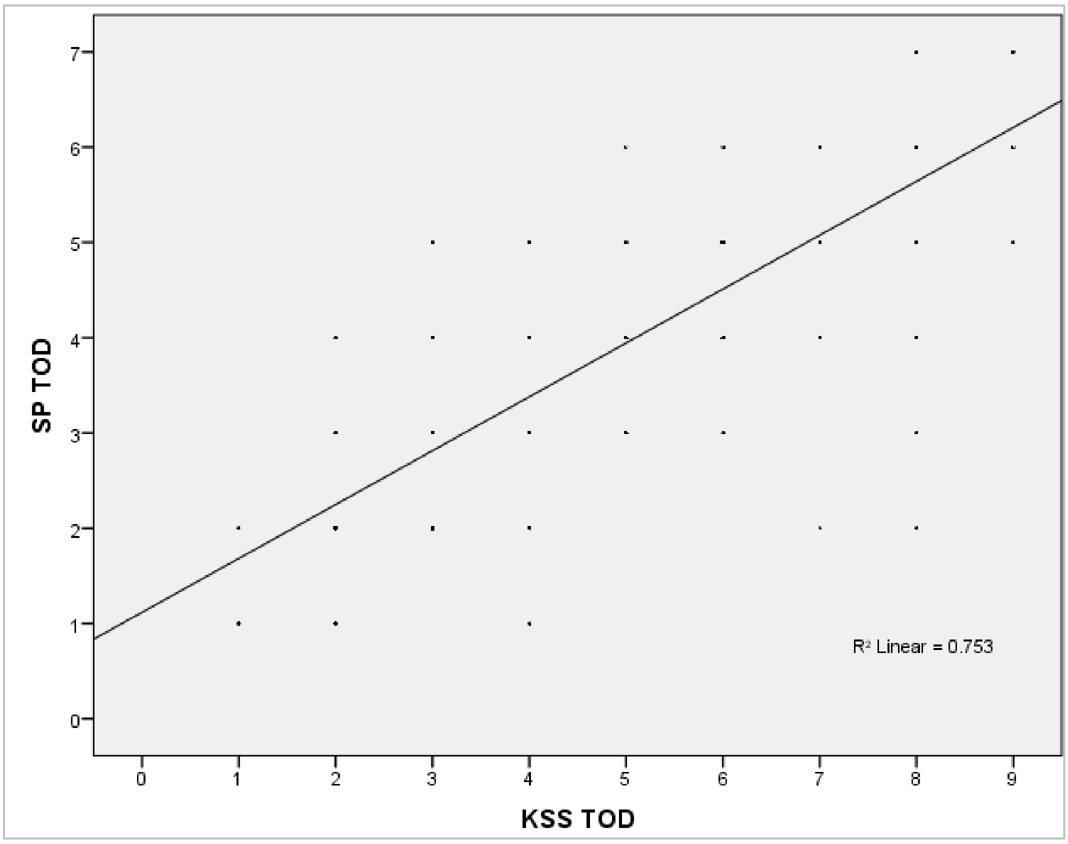


Figure 3-12*. Scatter plot showing KSS vs SP ratings at last Tod in FTL1 (1632 observations).*

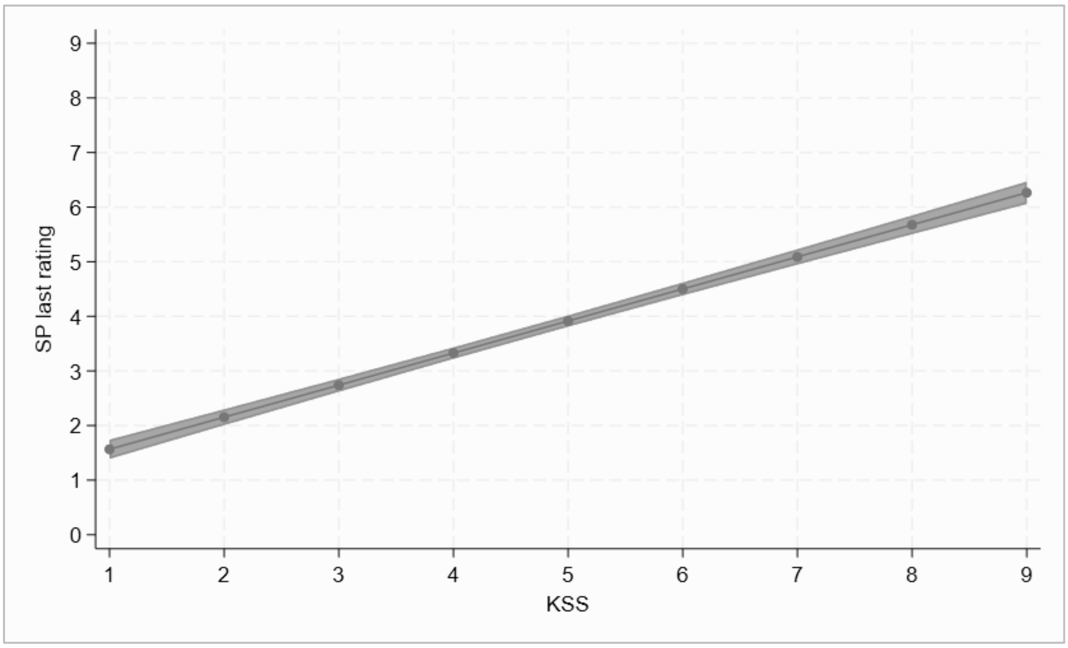


Figure 3-13. *Mixed model regression with KSS predicting SP values at TOD/last rating (Mean±se) based on the multi-sector data set of FTL2.0 (372 observations).*

3.1.3 PVT results

Figures 3-14 and 3-15 show the main results of the PVT performance in FTL1 (mean reaction time) and FTL2.0 (mean response speed). Neither of these results showed a significant difference between the FDP of interest and its control FDP(s).

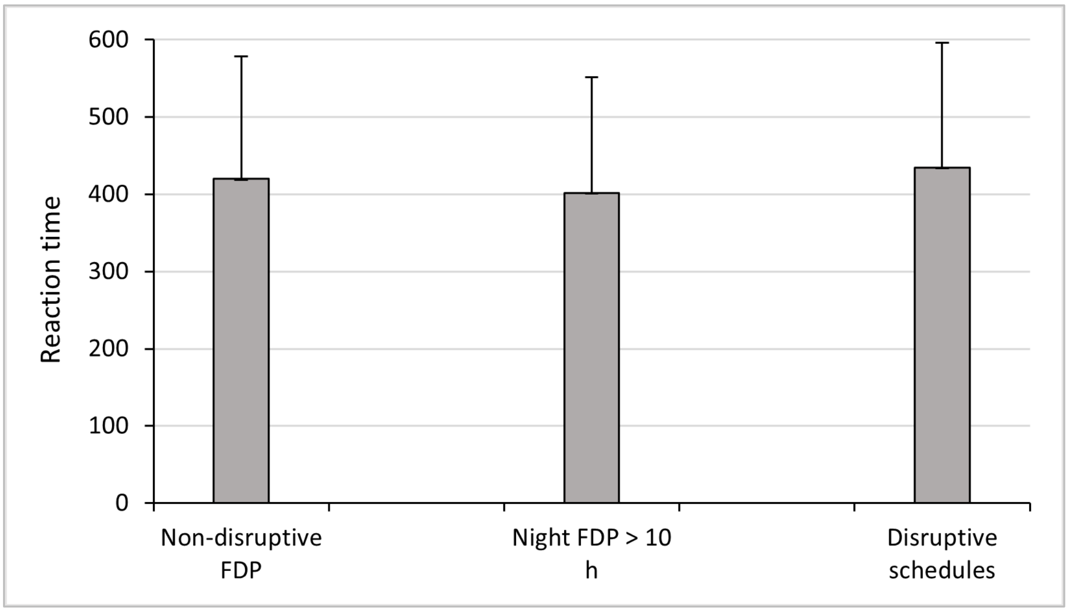


Figure 3-14. *Mean reaction time (in milliseconds) on the 5-min PVT performed before the last ToD of the FDPs in FTL1.*

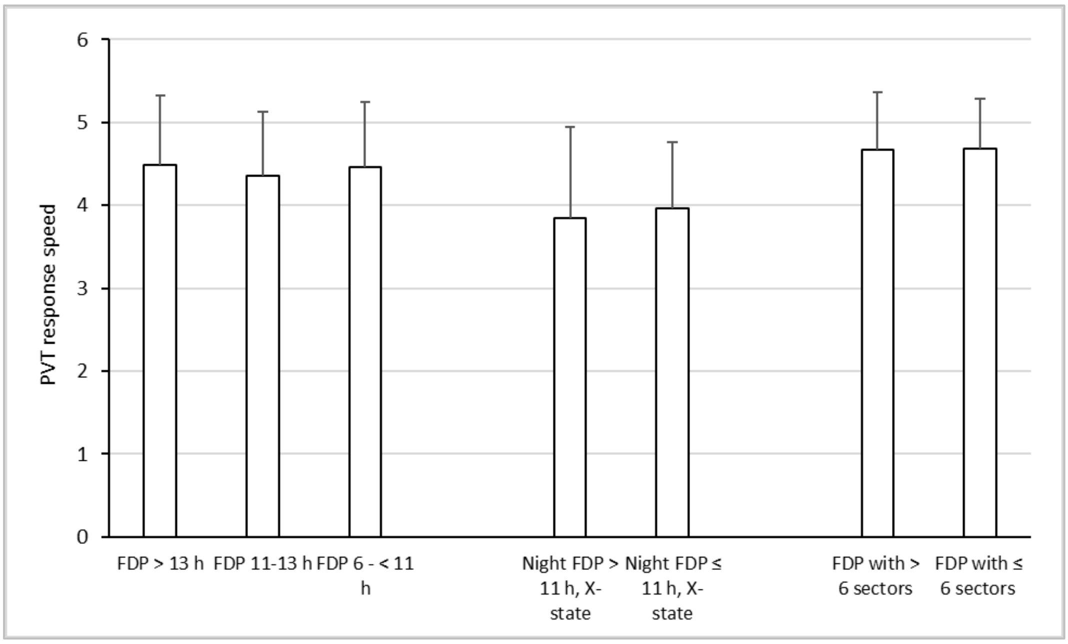


***Figure 3-15**. Mean response speed (mean 1/RT in seconds) on the 3-min PVT performed after the last Bon of the FDPs in FTL2.0.*

3.1.4 Predictors of high fatigue

Table 3-1 shows the main predictors of high fatigue (KSS ≥7) at last ToD by FDP of interest. These results are based on single- and multivariable regression analyses conducted in FTL1 and FTL2.0.

Of the FDP characteristics, longer duration was most often associated with high fatigue: it predicted high fatigue in the late finish FDPs, FDP >13 hour starting at the most favourable time of the day, FDP >11h flown in an unknown state of acclimatisation, and FDPs with >6 sectors in FTL2.0. An earlier start time predicted an increased probability (odds ratio) of high fatigue during the early start FDPs and encroachment on the WOCL predicted an increased probability (odds ratio) of high fatigue during the night FDPs when using daytime FDPs as a reference in FTL1.

In addition, shorter sleep and long prior awake time measured by a sleep log were significant predictors of high fatigue in night FDPs (shorter sleep), FDPs >13h starting between 06:00 - 13:29h (longer awake time), multi-sector FDPs (shorter sleep and longer awake time). Finally, crew category and individual factors were associated with high fatigue in night FDPs (crew category), and FDPs >11h flown in an unknown state of acclimatisation (crew category, age, gender).

The statistical results of single- and multivariable regression analyses on which the aforementioned results are based, are presented in more detail in FTL1 and FTL2.0 deliverables 2.3.

Table 3-1. Main predictors of high fatigue (KSS ≥7) at last ToD by FDP of interest[1].

FDP of interest	**Longer FDP duration**	**Earlier FDP start time**	**Later FDP end time**	**WOCL en-croach-ment**	**Higher number of sectors**	**Longer prior time awake**	**Shorter prior sleep time**	**Crew category (cabin vs flight)**	**Indi-vidual factors**
Early start FDP		X							
Late finish FDP	X								

Night FDP[2]				X			X	X	
FDP >13h, start 06:00-13:29h	X*					X*			
FDP >11h in X-state[3]	X							X	X (younger age; male)
FDP with >6 sectors	X*				X	X*	X		

[1]The identification of main predictors is based on datasets that include both the FDP of interest in question and its control FDP(s) (e.g., identification of predictors of high fatigue for FDPs with > 6 sectors is based on a dataset that include FDPs with one to nine sectors).

[2]Night FDP category includes also night FDPs >10h. In FTL1, no separate regression analysis was performed on night FDPs >10h.

[3]For FDPs > 11h in X-state, also having an inbound flight within an eastward rotation predicted high fatigue at last ToD. This FDP characteristic was only used in this particular FDP of interest.

**Significant in single variable analysis only.*

Grey shaded rows concern FDPs of interest studied in FTL2.0, the white rows concern FDPs of interest studied in FTL1.

3.2 Synopsis of the main results of controlled rest

To describe the main findings regarding the impact of controlled rest on aircrew alertness, and thus to determine if CR is an effective fatigue mitigation measure, the CR-related results of FTL1 and FTL2.0 were merged.

3.2.1 Results from FTL1

In the first FTL study, in-flight rest and CR were encountered in the hotspot survey preceding the measurement campaign. This survey was done in 2016 with the goal to determine in what types of FDPs high fatigue is most likely to occur. The results presented in FTL1 deliverable D2.1 (Identification of the potential fatigue hotspots) showed that CR is typically taken during non-augmented flights and is used by 8% of pilots in long-haul flights, while this proved to be 2% in both short-haul and regional flights. Even though CR was not measured in the measurement campaign of FTL1, naps during the non-augmented FDP types studied (long night flights and disruptive schedules) were reported. These naps (most probably controlled rest periods since all flights were non-augmented) proved to occur in 31% of the measured long night duties and in 14% of the other measured disruptive schedules.

3.2.2 Results from FTL2.0 per deliverable

3.2.2.1 Definition of Baseline (D1.1)

To determine the definition of baseline at the start of FTL2.0, a systematic literature review was performed in 2022. For CR, four medium- to high-relevance peer-reviewed articles were included. All four studies supported CR as a successful mitigation strategy for a sudden (unexpected) onset of fatigue.

Hilditch et al. (2020) found that CR was taken on 69% of the non-augmented flights studied in an EU-based airline, in which sleep was successfully achieved during 80% of these rest periods. CR was used most frequently on inbound flights (60%), night flights (55%), and flights of a certain duration <10h (63%).

When looking at the causes for CR, Gander et al. (1991) found in their US-based study that the timing at which CR is taken reflects the underlying cause for fatigue: CR taken early during the flight was likely due to accumulated sleep debt and/or adverse circadian phase (low body temperature), whereas CR taken later in the flight is more likely attributable to a longer time awake.

With respect to the effect of CR, Eriksen et al. (2006) found that, in their Scandinavia-based study, 2-pilot crew had higher levels of sleepiness during much of the flight, in particular 2–4 hours after Top of Climb. Their findings suggest that CR can be successfully used to mitigate fatigue in 2-pilot crews but does not lead to the same levels of alertness as 3-pilot crews using in-flight rest. Finally, Rosekind et al. (1994) randomly assigned pilots to nap in their seats during the cruise phase of flights between the US and Asia. As a result of this CR procedure, pilots in the nap group demonstrated significantly faster PVT response times and fewer lapses in attention during the final approach of the flight, while at the same time their subjective alertness ratings were not affected.

3.2.2.2 Definition of the target crew population (D1.2)

Since from both FTL1 and the literature review it was unclear how CR was used by flight crew (pilots), the airlines which were potentially eligible for participation in FTL2.0 were asked about the use of planned and unplanned rest. In addition, an online cross-sectional survey was distributed amongst their aircrew members to identify the circumstances under which CR is being used and to analyse its fatigue mitigating characteristics.

In total 16 airlines responded to the questions on the use of CR.

- Regarding the CR procedures, five airlines indicated not to have a CR procedure in place. These airlines were relatively small and flew primarily short-haul flights. Six airlines indicated to have specific CR procedures in their operations manual (A-OM). One airline indicated to explicitly recommend to use CR “when needed”, and another reported that CR could be used in a preventive way, thus during flights where high workload or fatigue is expected. Four other airlines referred to existing procedures such as those recommended by EASA, ICAO, and the Flight Safety Foundation.
- The airlines did not possess structurally collected data on CR usage. Indications regarding usage frequency ranged from “a few times per 1,000 legs”, “during at least 30% of the flights”, to “CR is used during almost all-night duties”. The airlines with CR procedures in place believed that CR is mainly used in (very) early and night duties that encroach the WOCL.

To gain a better understanding of the frequency and the circumstances under which CR is being used by aircrew members, a cross-sectional online survey was distributed within 14 airlines that were willing to participate as well. The total number of pilots who completed the survey and provided their informed consent, was 1,453. These pilots were categorised per haul-type of flights (regional, short-haul, medium-haul or long-haul) they flew most (>50% of total number flights during the past six months). In total, 81% (1,177) of the pilots who filled out the survey indicated to have used CR in the past three years. More than half of the participants (58%) indicated that they use CR to prevent and mitigate *expected* fatigue, while 26% indicated that CR is used to mitigate *unexpected* fatigue.

Pilots with experience in applying CR indicated that CR was used in 27% of their flights. On average, the participants answered that they slept during 66% of the CR occurrences. The mean duration of sleep was 24 minutes.

Moreover, 88% of the pilots indicated that they were convinced CR was effective in battling excessive fatigue. However, 12% of the participants reported that high fatigue was still present after taking CR. The participants also indicated that CR is taken most often in long-haul flights (>5 hours) during the night or early morning. Regarding the reasons to use CR, particularly the level of fatigue at the start of the flight, and the workload during the flight showed to be good indicators.

3.2.2.3 Findings from the analysis of the data collection campaign (D2.3)

Based on the data collection campaign conducted in FTL2.0, data from 102 flight crew members could be analysed. Of these participants, 39 (38%) reported using CR during at least one FDP (80 CR occurrences in total).

The results of the analyses showed that FDPs with CR within the FTL2.0 dataset were characterized by an increased number of time zones crossed, inbound flights during a westward rotation, and being in an unknown state of acclimatisation (X-state).

It was also found that CR was primarily used to mitigate sudden (unexpected) fatigue (68%), while in almost 30% of the observations CR was used for preventing expected fatigue. The average duration of CR was around 34 minutes. In 70 out of 80 of the CR observations (88%) pilots reported to sleep, with an average duration of about 24 minutes.

To examine the impact of CR on fatigue levels, a comparison was made between FDPs with CR, and matched FDPs of the same participants in which they did not use CR. The results showed that there was no significant difference in high fatigue levels at ToD of both types of FDPs. However, when looking at high fatigue levels across the entire FDP, higher fatigue was found for FDPs with CR in comparison with those without CR. This suggests that while CR may not completely eliminate fatigue, it could help to reduce the likelihood of high fatigue levels during critical phases of flight, such as at or after ToD.

3.2.3 Synopsis of results

The results of FTL1 suggested that CR was typically used more in long-haul flights than short- and medium-haul flights, and more during long night duties specifically compared to disruptive schedules. While the exact percentages are difficult to match due to the difference in nature of the data collection campaign, the FTL2.0 data does complement the results found in FTL1, and points in the same direction: CR is relatively frequently used in long overnight duties. Hence, while no further parallels can be drawn between the two campaigns, FTL2.0 uncovered more information on the reasons for, and effectiveness of CR, successfully using the same subjective metrics as applied to study the effects of the FDPs of interest on (high) fatigue (i.e. KSS and SP).

4. Discussion

This discussion section is divided into two subsections: one on safety performance metrics that can be used to assess the effectiveness of the current flight and duty time limitations and the other on the effectiveness of CR. These discussions are based on the synopsis of the results of FTL1 and FTL2.0 presented above.

4.1 Safety performance metrics for assessing the effectiveness of flight and duty time limitations

The results of FTL1 and FTL2.0 suggest that the KSS and the SP are suitable methods for assessing the current flight and duty time limitations, whereas the PVT (5-min and 3-min version) appears to be less suitable for this purpose in real-world aviation settings.

The primary focus of the FTL1 and FTL2.0 analysis was on KSS ratings at the last ToD of the FDP. Therefore, most of the evidence stem from these data. The main safety-relevant threshold derived from the KSS ratings at ToD was ≥7 (7-sleepy, but no effort to keep awake; 8-sleepy, some effort to keep awake; 9-very sleepy, great effort to keep awake), which was called “high fatigue” in both FTL1 and FTL2.0. On the SP scale the corresponding threshold was ≥5 (5-moderately tired, let down; 6-extremely tired, very difficult to concentrate; 7-completely exhausted, unable to function effectively).

With regard to what is an unacceptable KSS level, we refer to the methods section of deliverable D2.3 of the FTL2.0 study, where it is stated that a large proportion of participants drive off the road in a car simulator, or are taken off the (real) highway because of erratic (sleepy) driving at a KSS level of 7 (at level 8 or 9 virtually all participants are affected). Note that in both FTL1 and FTL2.0 we use KSS ≥7 as a cutoff (which also includes levels 8 and 9). From a safety point of view, KSS=7 would be unacceptable (as would KSS ≥7), particularly if this rating was reported repeatedly.

The results of FTL1 and FTL2.0 presented in this document show that KSS ratings of 7-9 reported at the last ToD of the FDP correlate with FDP characteristics such as the long duration, flying at night, starting early, finishing late, and the number of sectors. In addition, the state of acclimatisation[3] appears to be a factor in reporting KSS ratings of 7-9. Moreover, high KSS ratings appear useful not just for quantifying differences between the established FDP types but also within them. An illustrative example is the sub-categorization of night FDPs in FTL1 (see Figure 3-8). Also, the preliminary results of the breakdown of FDPs >13 hours by FDP type (late finish, night, non-disruptive) illustrates the potential of fatigue metrics based on self-ratings to examine the sources of fatigue in-detail (Figures 3-9a and 3-9b).

From a safety perspective, an alternative threshold for self-reported fatigue is KSS ≥8, which is called “very high fatigue” in FTL1 and FTL2.0. According to the results presented in the deliverables D2.3 of both FTL projects, the probability of reaching this level of fatigue is significantly lower than reaching KSS ≥7 during FDPs >13h starting between 06:00h and 13:29h (e.g. a probability of 0.40 for KSS ≥7 versus probability of 0.12 for KSS ≥8). The same holds for SP ratings of ≥5 and those of ≥6 during FDPs >13h starting between 06:00h and 13:29h (e.g. a probability of 0.52 for SP ≥5 versus a probability of 0.16 for SP≥ 6). Given the degree of difference between these two fatigue thresholds, it seems justified to use them both when assessing the effectiveness of the current flight and duty time limitations. KSS ≥8 and SP ≥6 appear to indicate fatigue that is characterised by

---

[3] defined on the basis of the time difference between reference time and local time and the time elapsed since reporting at reference time

severe self-reported difficulties staying awake (KSS ≥8) and concentrating and functioning (SP ≥6), whereas KSS ≥7 and SP ≥5 indicate fatigue that may not be characterised by such severe difficulties.

The PVT results were quite consistent across FTL1 and FTL2.0, showing no difference in the PVT outcome measures between the FDPs of interest and their control FDPs. The most apparent explanation for this finding is the context where the PVT was performed, that is, a naturalistic work environment including both external and internal distractors and other stimuli. Given this insensitivity of the PVT and the time needed to perform it, speak against its use for assessing the effectiveness of the current flight and duty time limitations in a real-world setting (as opposed to a controlled condition in e.g. a simulator). In addition, the lack of any established threshold to indicate high fatigue complicates interpreting the PVT results.

A somewhat different angle to safety performance metrics is to use more indirect outcome measures than those based on the measurement of fatigue and/or functional capacity. The results of FTL1 and FTL2.0 suggest that the amount of prior sleep and time awake measured by a sleep log are good candidates for this purpose. These outcomes predicted high self-rated fatigue at last ToD in more than one FDP of interest. Therefore, it can be assumed that requirements that lead to improvements in these outcomes (e.g. increase prior sleep) can also mitigate fatigue and related safety risks (Petrilli et al. 2006; Cosgrave et al. 2018).

4.2 Safety performance metrics for assessing the effectiveness of controlled rest

Although FTL1 did not specifically focus on CR, it was found that CR was used in 31% of the non-augmented long night duties and 14% of the other non-augmented disruptive schedules.

The FTL2.0 results confirmed that CR usage is frequently used in specific FDP’s. In the cross-sectional survey that was conducted before the data collection phase, 81% of the pilots reported to have used CR at least once in the past three years. The scientific literature showed that in some airlines CR is being used on 69% of non-augmented flights, although this high percentage may be caused by selection bias and airline-specific CR procedures. As the airlines included in the FTL2.0 data collection were more varied with respect to their CR procedures, a lower percentage (38%) of the participating flight crew reported CR during at least one FDP compared to findings from literature. Within the FTL2.0 dataset, FDPs with CR were characterized by the number of time zones being crossed, inbound flights during a westward rotation, and unknown state of acclimatisation.

The EASA guidance material (GM1 CAT.OP.MPA.210) mentions that CR is predominantly intended to manage sudden (unexpected) fatigue, but that it may also be used to manage fatigue that is expected to become more severe during higher workload periods later in the flight. The FTL2.0 cross-sectional survey showed that more than 50% of the participants use CR to mitigate expected fatigue. This finding might however be airline, FDP, and/or workload specific. Hence, the FTL2.0 field study showed that CR was primarily (in 68%) used to mitigate unexpected fatigue, while in 30% of the occurrences CR was used for the prevention of expected fatigue. The data gathered in FTL2.0 showed that on average, participants slept during 88% of CR occurrences, with a mean sleep duration of 24 minutes.

In summary, the findings highlight the importance of CR as a fatigue mitigation strategy for pilots, particularly during overnight long-haul flights that involve crossing multiple time zones. Although questions remain about the effectiveness of CR, previous scientific literature shows that CR is a successful strategy to mitigate unexpected fatigue (Eriksen et al., 2006; Rosekind et al., 1994). This is substantiated by the pilots in the FTL2.0 survey of whom 88% believed that CR is effective in mitigating excessive fatigue. The results of the FTL2.0 field study were somewhat less straightforward but also converge with the literature; CR could help to reduce high fatigue levels (as measured through KSS ≥7 ratings), during critical phases of the flight (such as ToD). Nevertheless, further research is needed to fully understand the effects of CR on fatigue levels and to optimize fatigue management strategies in aviation.

5. Conclusions

The conclusions drawn from the results of FTL1 and FTL2.0 are divided into two parts. The first part addresses the results as such, while the second part puts them into the context of assessing the effectiveness of the current flight and duty time limitations contained in sub-part FTL of Regulation (EU) 965/2012. All conclusions on the effectiveness of the current flight and duty time limitations as such and recommendations for their further development are presented in the deliverables D2.6 and D2.7 of the FTL2.0 study.

The following conclusions can be drawn from the results of FTL1 and FTL2.0:

- All disruptive schedules are associated with an increase in the probability of high fatigue (KSS ≥7) at last ToD as compared to daytime FDPs.
  - This increase is especially prominent in the context of late finishing and night FDPs (FTL1).
  - The most prominent increase in the probability of high fatigue appears to occur in the context of night FDPs that end at 06:00h or later (FTL1).
  - Also, FDPs >13 hours starting between 06:00h and 13:29h show a high probability of high fatigue at last ToD (FTL2.0). No evidence was found that FDP end time was a significant factor, i.e., that this result was due to late finishing or encroaching the night window.
- Longer FDP duration (FTL1 and FTL2.0) and a higher number of sectors (FTL1 and FTL2.0) are also associated with an increase in the probability of high fatigue at last ToD.
- The level of self-rated fatigue during FDPs assigned during other-than-airport standby periods appears to be associated more closely with the amount of prior sleep and the sleep-to-awake time ratio than the time spent awake per se (FTL2.0). However, this result is based largely on non-disruptive assignments only.
- CR is a commonly used fatigue mitigation strategy in flight crew members, particularly during overnight long-haul flights that involve crossing multiple time zones. Scientific literature showed that CR can contribute to mitigating unexpected fatigue during crucial phases of the flight, and the findings of FTL2.0 pointed in the same direction.

The following conclusion can be drawn from the results of FTL1 and FTL2.0 in the context of assessing the effectiveness of current flight and duty time limitations contained in sub-part FTL of Regulation (EU) 965/2012:

- The Karolinska Sleepiness Scale (KSS) can be used as a safety performance metric to assess roster-induced fatigue. Especially the probability (relative frequency) of KSS ≥7 and KSS ≥8 ratings is useful for this purpose. It is recommended to use both thresholds. KSS ≥7 may include ratings indicating increased fatigue without severe self-reported difficulties staying awake (KSS =7). In contrast, KSS ≥8 includes only ratings indicating increased fatigue with such self-reported difficulties.
- The Samn-Perelli Fatigue scale (SP) can also be used as a safety performance metric to assess roster-induced fatigue. Especially the probability (relative frequency) of SP ≥5 and SP ≥6 ratings is useful for this purpose. It is recommended to use both thresholds. SP ≥ 5 may include ratings indicating increased fatigue without severe self-reported difficulties concentrating or functioning (SP =5). In contrast, SP ≥6 includes only ratings indicating increased fatigue with such severe self-reported difficulties.
- PVT outcome measures (reaction time, response speed) seem to be insensitive to FDP characteristics such as start time, end time, duration, and number of sectors, and therefore cannot be considered as useful safety performance metrics for assessing fatigue mitigations in unsupervised conditions in real-world aviation.

- The amount of prior sleep and time awake appear to be useful outcome measures for indirectly assessing roster-induced fatigue and associated fatigue mitigations. Shorter sleep and longer awake time correlate well with a higher probability of high self-rated fatigue.
- Controlled rest may not completely eliminate fatigue, but it was found that CR is able to reduce the likelihood of high fatigue during critical phases of specific types of flights (overnight long-haul flights involving the crossing of multiple time zones), successfully using KSS ≥7 as fatigue threshold.

6. Bibliography

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; 23:240*-252.

Åkerstedt T., & Gillberg M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience; 52*:29-37.

Åkerstedt, T., D. Hallvig, 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.

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

Arthurs, M., Dominguez Veiga, J. J., & Ward, T. E. (2021). Accurate reaction times on smartphones: The challenges of developing a mobile psychomotor vigilance task. *2021 International Symposium on Wearable Computers*, 53–57.

Basner M, Mollicone D, Dinges DF. (2011). Validity and Sensitivity of a Brief Psychomotor Vigilance Test (PVT-B) to Total and Partial Sleep Deprivation. *Acta Astronaut. 69(11-12*):949-959.

Benderoth, S., Hörmann, H. J., Schießl, C., & Elmenhorst, E. M. (2021). Reliability and validity of a 3-min psychomotor vigilance task in assessing sensitivity to sleep loss and alcohol: fitness for duty in aviation and transportation. *Sleep, 44(11*), zsab151.

Cosgrave, J., Wu, L.J., van den Berg, M., Signal, T.L., & Gander, P.H. (2018). Sleep on Long Haul Layovers and Pilot Fatigue at the Start of the Next Duty Period. *Aerospace Med Hum Perform.*, 89, 19-25.

Elmenhorst, E. M., Rooney, D., Pennig, S., Vejvoda, M., & Wenzel, J. (2012). Validating a 3-min psychomotor vigilance task for sleep loss induced performance. *Journal of Sleep Research;21*: s1:115.

Eriksen, C. A., Åkerstedt, T., & Nilsson, J. P. (2006). Fatigue in trans-Atlantic airline operations: Diaries and actigraphy for two-vs. three-pilot crews. *Aviation, Space, and Environmental Medicine*, *77*(6), 605–612.

Gander, P. H., Graeber, R. C., Connell, L. J., & Gregory, K. B. (1991). *Crew factors in flight operations. 8: Factors influencing sleep timing and subjective sleep quality in commercial long-haul flight crews*. No. A-91106.

Grant, D. A., Honn, K. A., Layton, M. E., Riedy, S. M., and Van Dongen, H. P. A. (2017). 3-minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation. *Behav.Res. Methods* 49, 1020–1029.

Hilditch, C. J., Arsintescu, L., Gregory, K. B., & Flynn-Evans, E. E. (2020). Mitigating fatigue on the flight deck: How is controlled rest used in practice? *Chronobiology International*, *37*(9–10), 1483–1491.

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.

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, 1357-1362.

Powell, D., 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.

Rosekind, M. R., Graeber, R. C., Dinges, D. F., Connell, L. J., 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*. No. DOT/FAA/92/24.

Samn, S. W., & Perelli, L. P. (1982). Estimating aircrew fatigue: a technique with application to airlift operations. School of Aerospace Medicine Brooks Afb tx.

European Union Aviation Safety Agency

Konrad-Adenauer-Ufer 3
50668 Cologne
Germany

Mail EASA.research@easa.europa.eu

Web www.easa.europa.eu

An Agency of the European Union