European Commission

Effectiveness of Flight Time Limitation (FTL)

Final Report

MOVE/C2/2016-360
Classification: Public

Table of Contents

Executive summary 3
Abstract 8
Chapter 1: Introduction 9
Main objectives and scope 9
Specific objectives 9
Scope of the final report 10
Chapter 2: Baseline for the research 11
Review of the state of the art 11
Ranking of flight duty periods 13
Bio-mathematical modelling of flight duty periods 13
Online survey 13
The two duty periods ranked as most fatiguing 14
Chapter 3: Definition of the target population 15
Characterisation of the selected population 15
Chapter 4: Identification of potential fatigue hotspots 16
Online survey 16
Bio-mathematical modelling of roster data 17
Mapping the identified fatigue hotspots 18
Night FDPs of more than 10 hours 18
Disruptive schedule FDPs 19
Chapter 5: Definition of data collection scope and process 23
Study population 23
Methods 23
Procedures 25
Data handling 26
Chapter 6: Analysis of field data 27
Crew member data 27
Crew data representativeness and sample size 30
Data analysis: identifying fatigue hotspots 31
Main results 31
Fatigue at TOD during the FDPs of interest 31
Predicting high fatigue at TOD during the FDPs of interest 35
Additional analyses with reference to FDP duration 36
Additional results using alternative FDP categories 36
Summary of the results 39
Assessment of the strength of the evidence found 40
Discussion and conclusions 40
Understanding the main analyses results 40
Implications for fatigue mitigation 41
Need to revise regulations? 43
Chapter 7: Conclusions and recommendations 44
Conclusions 44
Recommendations 44
Follow-on actions 46
Chapter 8: Critique of the whole data collection activity 50
Critique assessment 50
References 51
Definitions 54
Abbreviations 55

Executive summary

EASA has been mandated to perform a continuous review of the effectiveness of the rules concerning flight and duty time limitations and rest requirements contained in Annexes II and III of Commission Regulation (EU) No 965/2012. The review commenced in 2017 with the commission of the current research.

The review set out to assess the impact on the alertness of aircrew of at least the following flight duty periods (FDPs):

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

According to the results of bio-mathematical model analyses and an online survey of aircrew, 'duties of more than 10 hours at the less favourable time of the day' and 'disruptive schedules' were the two duty periods ranked as the most fatiguing. The first phase of the research (described in this report) assessed the impact of these two FDPs on the alertness of aircrew. The remaining FDPs will be the subject of follow-up research.

Research approach

Data for the analyses were obtained from three sources:

- an online survey;
- rosters from airlines; and
- a field study.

The online survey was used to collect information from aircrew about perceived fatigue hotspots. Fatigue hotspots were defined as schedules that are associated with a high likelihood of high fatigue. The respondents could select from a list the 'fatigue items' that they deemed to be relevant causes of a fatigue hotspot. The total number of aircrew respondents was 15 680, of which 58.2 % were pilots and 41.8 % cabin crew members.

Data on worked rosters of pilots and cabin crew members, spanning approximately one year, were collected from the airlines participating in the study. To predict fatigue levels, the rosters were analysed using two bio-mathematical models: the BAM model and the SAFTE model. Overall, rosters from six airlines, representing a total of 264 746 FDPs, were analysed.

Volunteers could register to participate in the field data collection via an online portal. After registering, giving consent and familiarising themselves with the data collection app and measurement protocol, the volunteers started gathering data. The information collected concerned fatigue, alertness, mental effort and sleep for a period of 14 consecutive days. Data collection continued for a period of eight months, from July 2017 until February 2018. A total of 24 airlines participated in the data collection. These airlines were used as a reference set for the EU aviation sector. Overall, 381 crew members participated and 2 877 FDPs were analysed. The participating crew population consisted of 68 % pilots and 32 % cabin crew members.

Potential fatigue hotspots

The results of the online survey and the analysis of pilot and cabin crew roster data identified potential fatigue hotspots using bio-mathematical modelling. These hotspots provided a focus for the subsequent field data collection and its analysis.

Night duties of more than 10 hours

According to the bio-mathematical models, high fatigue scores occurred particularly in flight duties longer than 10 hours and that encroached partially or fully on the period between 2.00 and 4.59. The proportion of high fatigue was greater during night duties of more than 10 hours compared with the baseline data set containing all collected FDPs. This was confirmed by the survey results, as 'a long working day' was the most frequently indicated fatigue item and 'flying during hours when I would normally sleep' was the third most frequent: both items are linked to night duties of more than 10 hours.

Disruptive schedules

For *single* FDPs classed as 'disruptive schedule', we found a relatively high likelihood of high fatigue for late finishes and nights. This was underpinned by the survey results, as 'flying during hours when I would normally sleep' was the third most frequent item indicated as a cause of high fatigue. Conversely, prevalence of high fatigue was low for early starts. This finding, however, was somewhat countered by the survey results, as 'starting early' was the second most frequent relevant cause of high fatigue. The occurrence of high fatigue was greater for non-consecutive FDPs than for the baseline data set containing all the collected FDPs except early starts.

For the *consecutive* disruptive schedule FDPs (i.e. at least two in a row), the roster data set showed a relatively high prevalence of high fatigue for nights. For late finishes, the two bio-mathematical models showed different outcomes: high and very low prevalence. For early starts, the prevalence of high fatigue was very low. The occurrence of high fatigue in the roster data was greater for consecutive nights than for the FDP baseline set. The proportion of high fatigue was (effectively) zero for consecutive early starts. Results were inconclusive for consecutive late finishes.

Field data analysis

The primary field data analyses were performed using the Karolinska Sleepiness Scale (KSS) scored at top of descent (TOD) during the final sector of the FDP. A high level of fatigue was defined as scores ≥ 7 on the KSS ordinal scale. Numerous studies have indicated that performance levels start to decrease at KSS = 7.

The tables below provide a summary of the main results on high fatigue at TOD for the FDPs of interest.

Summary of results on high fatigue at TOD for non-consecutive FDPs

FDP of interest	***Main results on high fatigue[1] at TOD and its predictors***
Night duties (> 10h)	All night FDPs were associated with a high probability of high fatigue at TOD. Similar probability percentages were reported for short (≤ 10h) and long nights (> 10h).
Nights	The probability of high fatigue at TOD during night FDPs was higher than during daytime FDPs. Encroachment of the FDP on the WOCL and shorter prior sleep were the significant predictors. When night FDPs were analysed alone, shorter prior sleep explained the occurrence of high fatigue at TOD.
	To cover the continuum from evening to night, late finish plus night (start time before 24.00) FDPs were combined. A higher probability of high fatigue at TOD during these FDPs than during daytime FDPs was predicted by encroachment on the WOCL, shorter prior sleep, later FDP start time, and longer FDP duration. When late finish plus night FDPs were analysed alone, encroachment on the WOCL, earlier FDP end time, and shorter prior sleep explained the occurrence of high fatigue at TOD.
	To cover the continuum from late night to early morning, very early (3.00 – 4.59) and early (5.00 – 6.59) starting FDPs were combined[2]. A higher probability of high fatigue at TOD during these FDPs than in daytime FDPs was explained by earlier FDP start time and shorter prior sleep. When these FDPs were analysed alone, shorter prior sleep was the only factor that explained the occurrence of high fatigue at TOD.
	An alternative way of classifying FDPs was suggested. When applying this classification, deep early (start time 2.00 – 4.59) and early (start time 5.00 – 6.59) start FDPs[3] returned a similar probability of high fatigue at TOD. The highest probability of high fatigue at TOD was recorded for deep night FDPs that covered the entire night (start time 1.59 or earlier, end time 6.00 or later).
Early starts	The probability of high fatigue at TOD during early start FDPs was higher than during daytime FDPs. An earlier FDP start time was the only statistically significant predictor. When early start FDPs were analysed alone, none of the FDP-related characteristics explained the occurrence of high fatigue at TOD.
Late finishes	The probability of high fatigue at TOD was higher during late finish FDPs than during daytime FDPs. A longer FDP duration was the only significant predictor. When late finish FDPs were analysed alone, a longer FDP duration, an earlier FDP start time, and an earlier FDP end time explained the occurrence of high fatigue at TOD.

Summary of results on high fatigue at TOD for two consecutive FDPs

FDP of interest	***Main results on high fatigue[1] at TOD***
Consecutive early starts	Fatigue levels at TOD were similar for the first and second consecutive early start FDPs.
Consecutive late finishes	The probability of high fatigue at TOD appeared to be similar for the first and second consecutive late start FDPs.
Consecutive nights	Fatigue levels at TOD were similar for the first and second consecutive night FDPs.
Mix	The probability of high fatigue at TOD during mixes of disruptive schedule FDPs appeared to be greater than the corresponding probability in the baseline data set, containing all FDPs collected.

FDP (flight duty period). TOD (top of descent). WOCL (window of circadian low – the period between 2.00 and 5.59 in the time zone to which a crew member is acclimatised).

[1] A high level of fatigue was defined as Karolinska Sleepiness Scale (KSS) scores equal to or greater than 7 (= sleepy, but no effort to keep awake).
[2] FDPs starting between 3.00 and 4.59 are considered night FDPs in the current FTL, whereas those starting between 5.00 and 6.59 are not.
[3] The deep early FDPs are considered night FDPs in the current FTL, whereas early FDPs are not.

Conclusions and recommendations

Night FDPs, both longer and shorter than 10 hours, were associated with a high probability of reaching temporarily high fatigue at TOD (KSS ≥ 7). This is not fully reflected in the current FTL regulation and guidance material. The regulation and guidelines explicitly note the need for appropriate fatigue risk management and the importance of obtaining sufficient sleep in relation to night duties *longer* than 10 hours, but not for those *shorter* than 10 hours.

The research found that, within the FDPs defined as 'night' FDPs in the current regulations, three subgroups can be distinguished based on the probability of occurrence of high fatigue at TOD:

1. FDPs starting between 2.00 and 4.59;
2. FDPs ending between 2.00 and 5.59 and starting at 1.59 or earlier; and
3. FDPs ending at 6.00 or later and starting at 1.59 or earlier.

The current FTL does not include such a distinction; however, recognising these three subgroups could help operators to design effective fatigue risk management.

The factors that best predicted increased odds of high fatigue at TOD varied by FDP type. This suggests that fatigue mitigation measures should be based on various fatigue management strategies and tailored to the different FDP types and operator contexts.

The following recommendations are made in no particular order.

Recommendation 1
Within the FDPs that are defined as 'night' FDPs in the current regulation, three subgroups can be distinguished based on the probability of occurrence of high fatigue at TOD:

- FDPs starting between 2.00 and 4.59;
- FDPs ending between 2.00 and 5.59 and starting at 1.59 or earlier; and
- FDPs ending at 6.00 or later and starting at 1.59 or earlier.

It is recommended to include these subgroups in the definition of night FDPs to help operators to design effective fatigue risk management strategies.

Recommendation 2
The analysis provides evidence of high fatigue at TOD during late finish FDPs. It is recommended to require operators to apply appropriate fatigue risk management to mitigate the fatiguing effect of late finish FDPs, regardless of FDP duration.

Recommendation 3
The analysis provides evidence of high fatigue at TOD during both long duration (> 10h) and shorter duration (≤ 10h) night FDPs. It is recommended to require operators to apply appropriate fatigue risk management to mitigate the fatiguing effect of *all* night FDPs, regardless of FDP duration.

Recommendation 4
Within night FDPs, duty periods that end at 6.00 or later, combined with a start at 1.59 or earlier, show the greatest probability of high fatigue at TOD. It is recommended that the regulation define this category of FDP and require operators to pay specific attention to these FDPs when applying fatigue risk management for *all* night FDPs, as proposed in recommendation 3.

Recommendation 5

The analysis found shorter prior sleep to be a predictor of high fatigue at TOD for all night FDPs. The current guidance material for night duties (GM1 CS FTL.1.205) stipulates that it is 'critical for the crew member to obtain sufficient sleep' for night duties of more than 10 hours. It is recommended to amend the GM to state that it is critical for the crew member to obtain sufficient sleep before *all* night duties, regardless of FDP duration.

Recommendation 6

The analysis provides evidence of high fatigue at TOD during night FDPs. This phenomenon seems to be fairly independent of FDP characteristics (e.g. start and end times, duration), as long as the FDP in question meets the criteria for a night FDP. Prior sleep is the main predictor of eventual fatigue. We therefore recommend that for night FDPs, operators should be required to promote optimum use of sleep opportunities (i.e. before reporting and during the FDP).

How to implement recommendations 5 and 6?

Our research found that shorter sleep in the 24 hours prior to TOD is a predictor of high fatigue at TOD for all night FDPs. Ensuring that crew members obtain sufficient sleep is a shared responsibility of the operator and the crew. Current regulations (ORO.FTL.110 and 115) already describe the need for operators to provide resting opportunities and for crew to make optimum use of such opportunities. As this is essential for effective fatigue risk management, EASA and national civil aviation authorities may need to promote more decisively the use of resting opportunities. To implement this recommendation, we suggest addressing the practicalities conducive to attaining sleep prior to reporting or during the FDP. Providing rest facilities for crew members at or near the airport would improve the probability of obtaining sleep as close as possible to the start of the night duty (as referred to in GM1 CS FTL.1.205(a)(2) FDP on appropriate fatigue risk management for night duties). That might imply providing suitable accommodation (as defined in ORO.FTL.105 Definitions) at the reporting point for napping in the afternoon prior to a night duty, e.g. for commuting crew members and during the FDP when crew members are on the ground, such as during a long turnaround.

A way of improving opportunities for *in-flight* sleep is the use of an augmented crew (only applicable for longer flights). On augmented flights, the resting crew member uses a seat in the cabin (that meets regulatory requirements) or a separate rest facility.

As fatigue can sometimes occur unexpectedly, the use of controlled rest could be considered. 'Controlled rest' means an 'off-task' period of time that may include actual sleep. Our data shows that the napping frequency is higher than expected during night duties. According to GM1 CAT.OP.MPA.201, controlled rest is not proactive fatigue management (i.e. which is planned before the flight) and may be performed only to manage unexpected fatigue and to reduce the risk of fatigue during higher workload periods later in the flight.

We suggest promoting the development and use of controlled rest procedures to enable pilots and cabin crew to take a nap during night FDPs to manage unexpected fatigue and reduce the risk of fatigue during higher workload periods later in the flight. Pilots can take their controlled rest on the flight deck, whereas cabin crew would have to be provided with suitable bunks or seats by the operator. We also suggest that operators track the use of controlled rest, as it is a very useful indication of where additional more effective controls may be necessary.

Abstract

EASA has been mandated to perform a continuous review of the effectiveness of the rules concerning flight and duty time limitations and rest requirements contained in Annexes II and III of Commission Regulation (EU) No 965/2012. The review commenced in 2017 with the commission of the research reported in this document.

This first and current phase of the research assessed the impact of ‘night duties longer than 10 hours’ and ‘disruptive schedules’ on the alertness of aircrews. Under such circumstances, the research found an increased probability of high fatigue levels, especially during nights and duty periods with late finishes, among both pilots and cabin crew. The strongest predictors of high fatigue in these periods, compared with daytime duties, varied by type of flight duty. For early starts, the only significant predictor was the earlier start time itself. For late finishes, the only significant predictor was a longer duty duration. For nights, the pertinent predictors were encroachment on the window of circadian low (WOCL) and short prior sleep.

Based on the conclusions drawn from the outcomes of the analyses, six recommendations were made regarding further fatigue mitigation measures.

Chapter 1: Introduction

Main objectives and scope

The European Aviation Safety Agency (EASA) has been mandated to perform a continuous review of the effectiveness of the provisions concerning flight and duty time limitations and rest requirements contained in Annexes II and III of Commission Regulation (EU) No 965/2012.[1] The review commenced in 2017 with the commission of the current research.

In accordance with Article 9b of Regulation (EU) No 965/2012, the review set out to assess the impact on the alertness of aircrew of at least the following flight duty periods (FDPs):[2]

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

The review was broken down into smaller phases by the European Commission and EASA. Each phase focuses on specific FDPs. The first and current phase concerns the two types of duties expected to pose the highest level of fatigue.

Specific objectives

The current FTL research included the following work content:

- ranking of the chosen aircrew duty periods based on the expected level of fatigue and selection of the top two;
- identification of a representative population and a relevant type of operations to be used for data sampling purposes;
- detection of potential fatigue hotspots in commercial air transport (CAT);
- collection of objective and subjective data on mental effort, fatigue and performance from a target aircrew population;
- benchmarking of the research against other relevant sources;
- assessment of the fitness-for-purpose of the regulatory fatigue management controls to ascertain the need for any additional mitigations; and
- conclusions on the work performed and resulting recommendations.

[1] Commission Regulation (EU) No 965/2012 of 5 October 2012 laying down technical requirements and administrative procedures related to air operations pursuant to Regulation (EC) No 216/2008 of the European Parliament and of the Council.

[2] Duty period means a period which starts when a crew member is required by an operator to report for or to commence a duty and ends when that person is free from all duties, including post-flight duty (ORO.FTL.105 (11)).

Scope of the final report

This report presents the research performed in the current first phase of the FTL research. It provides an overview of the work conducted and summarises the following deliverables (Ds):

- D1 (Definition of the Baseline);
- D2 (Data Collection and Analysis):
  - D2.1 (Identification of Potential Fatigue Hotspots);
  - D2.2 (Definition of the Data Collection Process);
  - D2.3 (Performance of the Data Collection and Data Analysis);
- D3 (Conclusions and Recommendations):
  - D3.1 (Analysis of the Fitness for Purpose of the Current Safety Management Controls);
  - D3.2 (Benchmark of this Analysis with Other Relevant Sources); and
  - D3.3 (Conclusions and Recommendations).

Figure 1 illustrates the research flow. Each box represents a chapter in this report, starting from Chapter 2.

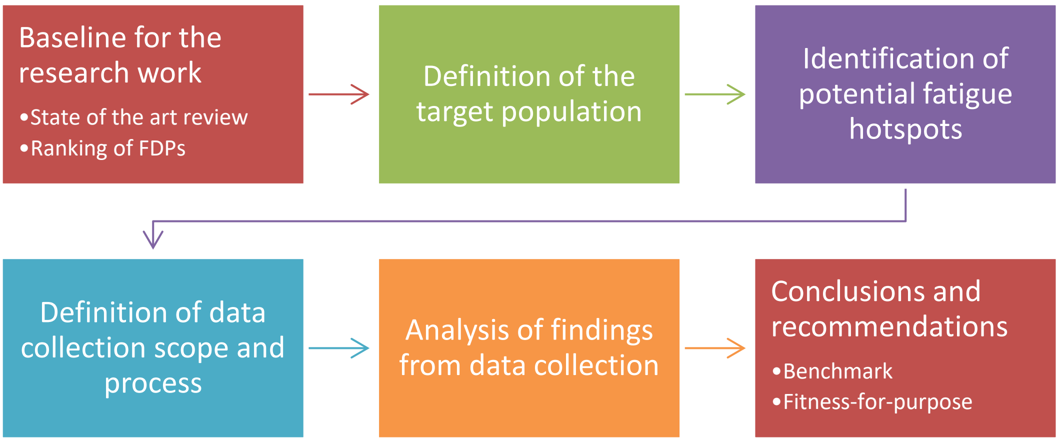

Figure 1 *Research flow chart*

Chapter 2: Baseline for the research

This chapter presents highlights of the review of the state of the art in aviation fatigue and alertness and ranks the FDP types by probability of high fatigue.[3]

Review of the state of the art

To review the state of the art in aviation fatigue and alertness, we compiled an overview of the existing literature, which included a number of studies that detailed similar research undertakings. These provided useful background information for the design of procedures and methods and for the subsequent data collection.

The publications examined were systematically assessed with emphasis on the following topics:

- selected population (*who was being measured?*);
- measurement techniques (*what was being measured?*);
- scale of data collection (*how many participants were being measured and for how long, in what geographical region?*);
- protocols followed (*when and how were measurements taken?*); and
- objectives and conclusions (*what was being studied and concluded?*).

The publications were categorised into two groups:

1. Fatigue-related research studies AND European commercial aviation (i.e. operations undertaken under EU regulations) AND published after 2006; and
2. Fatigue-related research studies AND non-European commercial aviation AND published after 2006.

The first group proved most relevant for the benchmarking exercise, as these publications were most closely aligned with the current research parameters. The 2006 cut-off publication date served to restrict the extent of the literature search to the current state of the art.

The literature search yielded 34 relevant publications, 10 European and 24 non-European studies.[4] Out of the 10 European studies, 3 showed a high degree of similarity with the current research as to research goals, target population, or measurements used.

1. Sallinen, M., Sihvolaa, M., Puttonena, S., Ketolac, K., Tuoric, A., Härmäa, M., Kecklundd, G., & Åkerstedt, T. (2017). Sleep, alertness and alertness management among commercial airline pilots on short-haul and long-haul flights. *Accident Analysis & Prevention*, *98*, 320-329.
2. Srivistava, A. S., & Barton, P. (2012). Collaboration on the human factors monitoring program (HFMP) study. NASA Report No TM-2012-216053.
3. Vejvoda, M., Elmenhorst, E. M., Pennig, S. B., Parh, G., Maass, H., Tritschler, K., Basner, M., & Aeschbach, D. (2014). Significance of time awake for predicting pilots' fatigue on short-haul flights: implications for flight duty time regulations. *Journal of Sleep Research*, *23(5)*, 564-567.

In addition to the literature review, we interviewed fatigue experts with recent experience in large-scale studies in an operational aviation environment. These

[3] This concerns an extract of D1 (Definition of the Baseline).
[4] References to all 34 publications are presented in the chapter References.

interviews offered additional information relating to aircrew fatigue and alertness field studies.

The literature review and interviews raised a number of issues that warranted special consideration for the design of the field data collection. Two areas of particular importance were the manner of inclusion and training of participants and the confidentiality of the data.

On inclusion and training of participants, three issues emerged:

- The key to successful recruitment is having all involved parties familiar with the reasons for, and in agreement with, the data collection. Recruitment involves collaboration between research teams, airlines, union representatives, and aircrew. Consulting with each party during protocol development has helped to facilitate study recruitment.
- Depending on the type of measures employed, training can be completed either in person or remotely using web-conferencing, or by providing online computer-based training. That last option is particularly relevant for large sample sizes and straightforward measurement protocols.
- A dedicated team should be formed for each participating airline to coordinate the data collection within the airline. The team should be composed of motivated personnel who are knowledgeable about the airline, as well as one or two representatives of the project consortium who are fully familiar with the measurement protocol. Participating crews should clearly know whom to address if problems arise or a lack of clarity is detected. When the sample size is somewhat smaller and the measurement protocol more complicated, the coordination team should provide the required training.

On confidentiality of data, too, several important issues emerged:

- Data on sleep/wake history and performance data is sensitive, as it may be used in the evaluation of workplace performance and in accident analysis and prevention. Researchers, employers, and participants often have concerns regarding the confidentiality of sleep and performance data.
- On the subject of protection of data confidentiality, the interviewees pointed to the need to build strong collaborative relationships with airlines, airline safety departments, unions, and regulators.
- The interviewees emphasised the need to provide full guarantees of the confidentiality of the gathered data. A number of interviewees said that they had used confidentiality agreements with the airlines for this purpose. Although relatively time consuming, these are sometimes necessary.

Ranking of flight duty periods

We ranked the six duty periods of interest according to the expected level of fatigue, based on objective and subjective estimates. For the objective estimates, bio-mathematical modelling was used. For the subjective estimates, an online survey was conducted. The current FTL research phase studied the two top ranked FDPs in terms of fatigue.

Bio-mathematical modelling of flight duty periods

Each of the six duty periods actually describes a range of possible specific schedules. To calculate fatigue levels with bio-mathematical models, we further refined the definitions and determined example duties. Consultation with EASA provided the following clarifications:

- ‘Most favourable time of day’ is intended to refer to operations between 08.00 and 21.59.
- ‘Least favourable time of day’ is intended to refer to operations that encroach on all or part of the night (i.e. the period between 2.00 and 4.59).
- ‘Disruptive schedule’ refers to consecutive early starts, late finishes, night duties, and combinations or a mix thereof.

This information was then used to create realistic schedules to be fed into the bio-mathematical models. The schedules were analysed using three bio-mathematical models:

1. Boeing Alertness Model (BAM, CrewAlert Pro 3.9.7);
2. Sleep, Activity, Fatigue, and Task Effectiveness, Fatigue Avoidance Scheduling Tool (SAFTE-FAST, v1.2.4.92); and
3. System for Aircrew Fatigue Evaluation (SAFE, v7.0).

Online survey

A survey was developed in a number of iterations to ensure high-quality questions using a language and format that was easy for respondents to understand. The survey asked aircrew members to assign a fatigue rating to each of the six duty periods based on their experience of the three previous years. The 9-point Karolinska Sleepiness Scale KSS was used for the ratings, which were then ranked on this basis.

Experts, such as researchers, safety experts and schedulers, were presented with a slightly different version of the survey, as they were not assumed to have own experience of the different duty periods. They were asked to rank the duty periods on a 6-point scale based on their expertise, with 1 being the most fatiguing. This ranking was then transformed into a rating system comparable to that used by the aircrews.

The aircrew responses were filtered, and only responses from aircrew members either working for a European airline or not currently working but living in Europe, were included in the analysis. All expert responses were included, irrespective of where the expert resided. When the survey was filled more than once from the same IP address, responses were checked for similarities. In cases of high resemblance, only a single stream of survey output was used in the data analysis.

The final survey data set consisted of 15 806 respondents: 51.1 % were pilots, 48.1 % cabin crew members, and 0.8 % held another occupation, such as researcher, safety expert, or scheduler.

The two duty periods ranked as most fatiguing

Based on the findings from the bio-mathematical modelling and the survey, FDPs were ranked as follows:

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

All of the models ranked 'duties of more than 10 hours at the less favourable time of the day' in their top two. For 'disruptive schedules' (i.e. early starts, late finishes, and nights) the ranking results differed per model. However, at least one of the disruptive duties was ranked in the top three of each model. This was supported by the rankings produced based on the survey of researchers, safety experts, and schedulers.

This led to the conclusion that 'duties of more than 10 hours at the least favourable time of the day' and 'disruptive schedules' are the two most fatiguing duty periods of the six FDPs of interest. These two FDPs therefore became the focus of the first and current research phase.

Chapter 3: Definition of the target population

This chapter describes the crew member population and the rationale that justifies the selection.[5]

Characterisation of the selected population

The identification of a representative population and the relevant types of operations to be used for data collection purposes followed a two-step process:

- establishment of a subset of Member States that are representative of the existing conditions in the EU aviation sector as a whole; and
- definition of the criteria for a screening of commercial air transport (CAT) aeroplane operators and subsequent execution of the screening to achieve a representative mapping of air operations.

Several criteria, which are listed below, were taken into account in the screening of the CAT operators.

- Volume of air operations (as a function of the number of aircraft), as this was considered a key determinant of operators' exposure to fatigue. An internet search was performed to this end.
- The extent to which operators used deviations or derogations from the EU FTL Regulation. This was based on information on exemptions and derogations from the EASA website[6]. Airline operators that used such flexibility were excluded.
- The type of FDPs performed by the operators. This information was gathered through expert opinion.

These criteria resulted in the following set of candidate EU CAT operators. These candidates were approached and asked to participate in the data collection. Any other CAT operator could also volunteer to participate.

Adria Airways, Aegean Airlines, Air Baltic, Air Berlin, Air Europa, Air Nostrum, Alitalia, ASL Airlines Belgium, BRA Braathens, British Airways, Cargolux, Condor, Croatia Airlines, Czech Airlines, Flybe, Iberia, Icelandair, KLM, LOT Polish Airlines, Lufthansa, Lufthansa Cargo, Norwegian Air Int., Norwegian Air Shuttle, Ryanair, Scandinavian Airlines, Smartwings, TAP Portugal, TAROM, Thomas Cook Scandinavian, Vueling, WIZZ Air, and WOW Air.

[5] This is an extract of D2.2 (Definition of the Data Collection Process).
[6] Https://www.easa.europa.eu/document-library/regulations/flexibility-provisions.

Chapter 4: Identification of potential fatigue hotspots

This chapter reports the results of the work performed to identify potential fatigue hotspots among the target population. These were determined using both an online cross-European survey and bio-mathematical modelling of pilot and cabin crew roster data.[7] Fatigue hotspots were defined as schedules that are associated with a high likelihood of high fatigue. The fatigue hotspots that were identified by aircrew members provided the focus for our data collection and analysis.

Online survey

We developed and used an online survey to collect aircrew insights about perceived fatigue hotspots. That same survey was used as an information source for the subjective ranking of the six duty periods described in 'Chapter 2: Baseline for the research'. The respondents could select from a list the 'fatigue items' that they deemed to be relevant causes of a fatigue hotspot. They could also describe in their own words (i.e. answering open questions) how the rosters affected their fatigue, when they felt most fatigued during their duty, and what conditions worsened their fatigue.

The total number of aircrew respondents was 15 680, or some 10.6 % of the entire aircrew population base in Europe. Of these, 58.2 % were pilots and 41.8 % were cabin crew members. Respectively, 22.0 % and 7.5 % of the entire European pilot and cabin crew population participated. Figure 2 illustrates the geographic distribution of survey respondents (left) and the corresponding proportions for the entire population base (right).

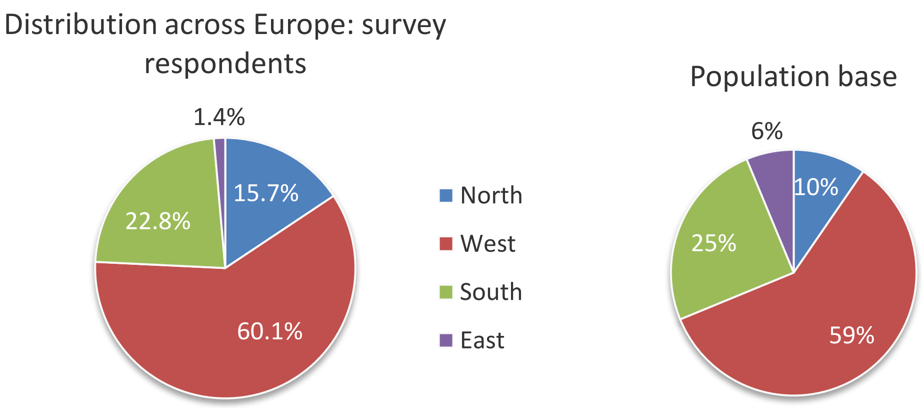

Figure 2 *Geographical distribution across Europe of survey respondents and aircrew population. The total aircrew population (right) comprises all aircrew members working for European CAT operators. The size of this population was estimated by the authors based on EASA statistics, internet research and random checks at airlines.*

The mean age of pilots (4.5 % female) was 42 years and 4 months; for cabin crew (61.5 % female), it was 40 years and 10 months. Of the crew respondents, 27.5 % indicated that they worked for a point-to-point operator, 61 % for a network operator, 3.3 % for a cargo operator, and 8.2 % for another type of airline.

[7] This is an extract of D2.1 (identification of Potential Fatigue Hotspots).

Bio-mathematical modelling of roster data

Data on worked rosters of pilots and cabin crew members, spanning approximately one year, were collected from six airlines. In total, rosters representing 264 746 FDPs were analysed. The months of July to October showed the highest numbers of FDPs in the roster data set. Looking at the geographical distribution and types of operations included in the data set, a lack of rosters from the northern region of Europe was observed.

To predict potential fatigue levels, the airline rosters were analysed using two bio-mathematical models: the BAM model and the SAFTE model. BAM predicts alertness with outputs expressed using the Common Alertness Scale (CAS). CAS values range from 0 (least alert state) to 10 000 (most alert state). CAS values were linearly mapped against the KSS values[8], with a KSS score of 9 (very sleepy, great effort to keep awake, fighting sleep) mapped to a CAS score of 0 and a KSS value of 1 (extremely alert) mapped to a CAS score of 10 000. For each pilot and cabin crew member, BAM was configured to assign a single alertness prediction at top of descent (TOD). TOD was defined as a half hour before wheels on ground.

The SAFTE model provides a percentage of performance effectiveness (Effect) from 0 (low effectiveness) to 100 (high effectiveness). There is an inverse relation between the SAFTE effectiveness scale and the KSS and Samn-Perelli (SP) scale[9]. A SAFTE value of 20 corresponds to a KSS value of 9 (very sleepy, great effort to keep awake, fighting sleep), and a SAFTE value of 100 corresponds to a KSS value of 1 (extremely alert). As to the SP scale, a SAFTE value of 20 corresponds to SP 7 (completely exhausted, unable to function effectively), and a SAFTE value of 100 corresponds to SP 1 (fully alert, wide awake). For each pilot and cabin crew member, the SAFTE model was configured to assign an Effect prediction at TOD.

We performed the analysis using as dependent variables CAS and Effect estimated at TOD during the final sector of the FDP. Data analysis consisted of two steps.

Step 1: Check for predicted high fatigue scores

The first step in data analysis was to check for predicted high fatigue scores. The goal here was to identify whether predicted high fatigue scores occurred in night FDPs of more than 10 hours and FDPs with disruptive schedules.

The following values were taken to define a high predicted fatigue level: BAM CAS scores equal to or below 2 500 and SAFTE Effect values equal to or below 77, i.e. equivalent to KSS scores of 7 or higher, and 6 or higher on the SP scale. A high predicted fatigue level was also defined by an Effect value lower than 88.5 for a minimum duration of 90 minutes (referred to as TimeLowEffect), i.e. equivalent to a KSS score of 5 or higher, and a score of 4 or higher on the SO scale, both for a minimum duration of 90 minutes.

[8] KSS is a 9-point 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 difficulty remaining awake; 8. Sleepy, some effort to keep alert; 9. Very sleepy, great effort to keep awake, fighting sleep.

[9] SP is a 7-point scale: 1. Fully alert, wide awake; 2. Very lively, but not at a 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.

Step 2: Find clusters of variables

In the second step, multiple regression models were developed to determine the characteristics of FDPs for which high fatigue levels were predicted by the models.

FDP-related characteristics that may contribute to fatigue were defined based upon the following sources:

- the online survey findings;
- parameters in the bio-mathematical models that were used for analysing the roster data;
- scientific literature review; and
- ideas and suggestions from the scientific committee and consortium members.

Mapping the identified fatigue hotspots

Below we present a map of the identified fatigue hotspots, based on the findings from our bio-mathematical modelling of the roster data. The survey among pilots and cabin crew members served mainly to help us interpret the results from the bio-mathematical models.

Night FDPs of more than 10 hours

Check for predicted high fatigue scores

We sought to assess the prevalence of predicted high fatigue scores during duties of more than 10 hours encroaching on the night. Our results show that, according to the bio-mathematical models used to examine the roster data, high fatigue scores did occur in flight duties longer than 10 hours that encroached partially or fully on the period between 2.00 and 4.59. Moreover, the proportion of high fatigue was greater in these FDPs than in the baseline set containing all the collected FDPs. This was confirmed by the survey results, as 'a long working day' was the most frequently indicated fatigue item and 'flying during hours when I would normally sleep' as the third most frequent, both of which are linked to night duties of more than 10 hours.

Find clusters of variables

We sought clusters of FDP-related characteristics that might impact on the predicted fatigue level during night FDPs longer than 10 hours. The results from the multiple regression models, using either CAS or Effect values as the dependent variable, differed to some extent. This can only be the result of differences between the bio-mathematical models used, as the same data sets served as inputs to the analyses. Although both models are 'two-process models', they differ in the way they represent and implement the two processes. Their mathematical representation of each process is different, the relative weighting of the two processes is different, and the manner in which the two models estimate a pattern of sleep associated with a sequence of duties is also different. Moreover, the two models are rooted in different types of research data (validated primarily against either alertness ratings or cognitive performance). This may account for the discrepancies in the models' weightings of fatigue factors.

The clusters of variables associated with night FDPs of more than 10 hours included the following predictors:

- sleep prediction in the 24 hours prior to TOD of the final sector;
- start and end time of the FDP;
- time in window of circadian low (WOCL[10]; between 2.00 and 4.59); and

[10] WOCL refers to the period between 2.00 and 5.59 in the time zone to which a crew member is acclimatised.

- number of time zones crossed eastwards and westwards.

The predicted duration of sleep in the 24 hours preceding the FDP was included in the multiple regression model because sleep is the primary recovery mechanism for fatigue. The bio-mathematical models differ in the way they estimate sleep; as an example, CAS includes no consideration of the effect of sleep on the odds of high fatigue. There is, however, an indirect link, since survey respondents most frequently indicated 'insufficient time between duties' as a contributing factor to fatigue. Insufficient time between duties may be interpreted as not enough time to get a good sleep.

For night FDPs of more than 10 hours, the earlier start and later end time of the duty period were included in the multiple regression models. This was not surprising, as the body's circadian rhythm has a major effect on fatigue levels. This is confirmed by the significance of the WOCL variable. Here, we see a clear link with the survey results, as the most frequently indicated period when crew were most fatigued was 'in the WOCL', and the second 'at the end of the duty'. Note that in the survey, fatigue was not reported specifically at TOD, contrary to what was observed in the roster data analyses.

For night FDPs longer than 10 hours, the number of time zone crossings was included in the models, because crossing more time zones results in longer FDPs that potentially encroach (part of) the WOCL. There was only a limited reference to time zone crossings in the survey responses, i.e. in- and outbound flights crossing more than six time zones was the least frequently indicated fatigue item.

Relevant rules in FTL

The predictors included in the multiple regression model for flight duties longer than 10 hours that encroached partially or fully on the night are linked directly or indirectly to the following rules in the flight duty time limitations and rest requirements (Commission Regulation (EU) No 83/2014):

- start of the FDP at reference time (ORO.FTL.205 FDP).
- sleep opportunity in 24 and 48 hours (ORO.FTL.205 FDP and ORO.FTL.235 Rest Periods);
- flight duration (ORO.FTL.205 FDP and ORO.FTL.210 Flight Times and Duty Periods);
- number of time zones crossed (ORO.FTL.235 Rest Period); and
- duty time in WOCL (ORO.FTL.205 FDP).

Disruptive schedule FDPs

Check for predicted high fatigue scores

We sought to assess the prevalence of predicted high fatigue scores during the different types of consecutive or non-consecutive disruptive flight duties. According to the bio-mathematical models used to examine the roster data, predicted high fatigue scores did occur for most types of disruptive schedules, but there were some differences between the different types of schedules.

For the *non-consecutive* disruptive FDPs, we found a relatively high prevalence of fatigue in late finishes and nights. This was underpinned by the survey results, as 'flying during hours when I would normally sleep' was the third most frequent item indicated as a cause of high fatigue. Conversely, prevalence of high fatigue was low for early starts. Somewhat contradictorily, 'starting early' was the second most frequent item indicated as a relevant cause of high fatigue. However, it should be noted that the survey questions used 'fatigue' as a broad term including physical fatigue, mental fatigue and sleepiness, and that the survey questions were not specifically focused on TOD. Nonetheless, the proportion of high fatigue was greater for the non-consecutive disruptive FDPs than for the baseline set, except for early starts, and the proportion for late finishes was just below the 1.0 relative ratio for Effect.

For the *consecutive* disruptive schedules (i.e. at least two in a row), we found a relatively high prevalence of fatigue for nights. The two bio-mathematical models showed different outcomes for late finishes: high prevalence for CAS and very low prevalence for Effect. For early starts, the prevalence of fatigue was very low. The proportion of high fatigue was greater for consecutive night disruptive schedules than in the FDP baseline set. The proportion of high fatigue was (effectively) zero for consecutive early starts, and results were inconclusive for consecutive late finishes, that is to say that the proportion was larger than 1 for CAS, and zero for Effect.

Note that the same FDP baseline set, with the same implications, was used for disruptive flight schedules as well as for night FDPs longer than 10 hours.

Find clusters of variables

We sought clusters of FDP-related characteristics that might impact on the level of predicted fatigue during *consecutive* disruptive flight duties. The resulting multiple regression models per consecutive disruptive duty differed to some extent, due to the differences between the bio-mathematical models used.

Consecutive early starts

Multiple regression models for consecutive early starts could not be computed because the prevalence of predicted high fatigue was very low, i.e. 1 (CAS) and 0 (Effect) in 6 895 observations. 'Consecutive early starts' were ranked as the seventh (out of 11) most frequent answer to describe how the preceding roster affected fatigue.

Consecutive late finishes

Only for the CAS measure could a multiple logistic regression model be developed, as there were no valid cases for modelling of the Effect measure. The resulting model for consecutive late finishes included the following predictors:

- late or very late finishes (end of duty between 23.00 and 1.59);
- sleep opportunity in darkness in the 48 hours prior to TOD of the final sector;
- FDP duration; and
- time awake prior to TOD.

The predictor 'late or very late finish' was associated with a relatively strong increase in the odds of high fatigue. Survey respondents indicated 'finishing late' as the seventh (out of 14) most frequent fatigue item.

The duration of a sleep opportunity in darkness in the 48 hours preceding the FDP was included in the multiple regression model, because sleep is the primary recovery mechanism for fatigue. Here, the survey results provide an indirect link, as 'insufficient time between duties' was the most frequently indicated contributing factor to how the preceding roster affected fatigue.

With regard to FDP duration, too, the survey results provide a link. Respondents indicated 'long working days' as the third most frequent contributor to fatigue. This was confirmed in the model by the significance of an extended time of wakefulness prior to TOD. 'At the end of the duty' was the second most frequently answer to the question 'when do you feel most fatigued?'.

Consecutive night flights

The multiple regression models for consecutive nights included the following predictors:

- time in WOCL;
- start and end of FDP; and
- FDP duration.

Sleep opportunity (or prediction) did not emerge as a strong predictor in the multiple regression models. Another recovery-related predictor that was not included is rest period. These refer to the period directly prior to the FDP.

WOCL was identified in the model as a predictor for the Effect measure. The importance of this variable also appeared in the survey results. In response to the question 'when do you feel most fatigued?', respondents indicated 'in the WOCL' most frequently.

For consecutive night flights, a correlation with fatigue at TOD was found for the start and end of the FDP at the reference time. This is similar to the results for night FDPs longer than 10 hours and is aligned with previous studies. The physiological mechanism underlying this phenomenon is well known: fatigue during night FDPs is due to the circadian downswing of alertness and extended time awake; in addition, the night sleep may be short, especially before inbound night flights when the local time and the biological clock are misaligned, causing circadian disruption.

Survey respondents indicated 'a long working day' as the third most frequent contributor to high fatigue in disruptive flight schedules. This variable (i.e. FDP duration) was also included in the multiple regression models and appears to be related to the end and start time of the FDP.

Mix of early starts, late finishes, and night flights

The multiple regression models for a mix of disruptive schedules included the following predictors:

- early starts (start of duty between 5.00 and 5.59/6.59 – depending on early or late type) and late or very late finishes (end of duty between 23.00 and 1.59);
- transitions from late finish to night;
- time in WOCL;
- start and end of FDP; and
- time of day.

The predictors 'early starts' and 'late or very late finishes' were included in the CAS model. These variables overlapped with the grouping variable for mix of disruptive schedules. Given the grouping variable, it also makes sense that several WOCL variables were included as predictors. The same applies to transitions from late finish to night duties.

Sleep was not identified as a predictor in either model. However, later FDP start and end times and a later time of day did increase the odds of high fatigue.

Relevant rules in FTL

The relevant predictors for different consecutive disruptive schedules are linked directly or indirectly to the following rules in the flight duty time limitations and rest requirements (Commission Regulation (EU) No 83/2014):

- start of the FDP at reference time (ORO.FTL.205 FDP);
- sleep opportunity in 24 and 48 hours (ORO.FTL.205 FDP and ORO.FTL.235 Rest Periods);
- flight duration (ORO.FTL.205 FDP and ORO.FTL.210 Flight Times and Duty Periods);
- number of time zones crossed (ORO.FTL.235 Rest Period);
- duty time in WOCL (ORO.FTL.205 FDP); and
- rest period provided before undertaking an FDP (ORO.FTL.235 Rest Period).

Chapter 5: Definition of data collection scope and process

This chapter details the scope and process of the data collection.[11] The resulting research protocol was reviewed by the Dutch ethics review committee on research involving human subjects ('Medisch Ethische Toetsing Commissie') at the Amsterdam UMC, reference W17_117.136.

Study population

Candidate CAT operators were first approached and asked to participate in the data collection (see also 'Chapter 3: Definition of the target population'). Crew members were then recruited for the field study from within the set of CAT operators that had agreed to take part. To be eligible to participate in this study, participants had to meet the following criteria:

- be a pilot or cabin crew member working for one of the EU CAT operators taking part in the research;
- operate (non-augmented; without in-flight rest) one or both of the earmarked duty periods in the foreseen period of data collection:
    - FDPs of more than 10 hours at the less favourable time of the day;
    - disruptive schedules.

Methods

We collected crew data on fatigue, alertness, mental effort and sleep, and logged it using the CrewAlert app[12]. A selection of participants was also asked to wear an actigraph, a device to monitor activity levels.

Fatigue

Field studies often use sleepiness or fatigue ratings. One frequently used measure is the KSS, which measures sleepiness on a 9-point scale from 'extremely alert' to 'very sleepy, great effort to keep awake, fighting sleep'. KSS has been validated and is used to measure subjective sleepiness in both laboratory and field studies[13]. Another frequently used rating scale is the Samn-Perelli (SP) crew status check. The SP scale measures fatigue on a 7-point scale, with scores ranging from 1 ('fully alert, wide awake') to 7 ('completely exhausted, unable to function effectively'). The SP scale was developed specifically for use by flight crew[14]. It has been used in studies on sleep loss, fatigue, and performance of flight crew[15], as well as in laboratory studies[16]. Both KSS and SP were used as measures in the data collection.

Performance

The psychomotor vigilance task (PVT) is a widely used and validated performance measure[17]. The PVT is a sustained-attention, reaction-timed task that determines the speed with which subjects respond to a visual stimulus. This study used the 5-minute version. Note that only pilots were asked to complete the PVT. Cabin crew were not

[11] This is an extract of D2.2 (Definition of the Data Collection Process).
[12] The CrewAlert app that was used for data collection only runs on iOS devices.
[13] Gillberg et al. 1994; Härmä et al. 2002; Kaida et al. 2006
[14] Samn & Perelli, 1982
[15] Samel et al. 1997
[16] Fergusin et al. 2012
[17] Loh et al. 2004; Basner et al. 2011

asked to complete the PVT because, in contrast to the flight deck, it is difficult to find a quiet place in the cabin to do the PVT without interruption.

Sleep

Actigraphy is an objective, non-intrusive and valid measure of sleep quantity and timing. Total sleep time measured by actigraphy is highly correlated with that measured by polysomnography among flight crews while in flight and during layovers[18]. This study used the Philips Respironics Actiwatch 2, a device that has shown high similarities with the sleep measured by polysomnography[19]. Due to the restricted availability of actigraphs, only a subgroup of participants within each airline was asked to wear a device.

A sleep diary or sleep log is a record of an individual's sleeping and waking times. This study used the CrewAlert app for sleep logging.

Mental effort

Mental effort was measured by means of a subjective rating scale. Using an app, participants moved a sliding bar to indicate the level of mental effort they expended in a particular duty period. The bar ran from ‘almost no effort’ to ‘extreme effort’ and was based on the Rating Scale Mental Effort (RSME)[20].

Work-related inputs

We asked aircrew to collect the following work-related information:

- duty period: departure and arrival times at airports;
- number of legs per duty period;
- time zone shift: the time zone at duty start and end;
- briefing and debriefing duration;
- in-flight sleep duration: the amount of sleep obtained during flight duty;
- hassle factors: daily bothers selected from a predefined list based on Vejvoda et al. (2014) research on pilots, supplemented with items added for cabin crew:
  - no break;
  - bad weather;
  - demanding airport;
  - high-density airspace;
  - sluggish ground handling;
  - tight crew rotation;
  - technical defect;
  - duty change on short notice;
  - hotel (noisy/low quality);
  - difficult passengers;
  - low quality food;
  - long travel from airport to hotel;
  - short turnaround;
  - delay/time pressure;
  - emergency;
  - critical fuel status;
  - abnormal procedures;
  - waiting times between flights;
  - difficulty getting through security;
  - insufficient baggage handling; and
  - other.

[18] Signal et al. 2005
[19] Kushida et al. 2001; Edinger et al. 2004
[20] Zijlstra, 1993

Demographics
We asked crew members to provide the following demographic information: gender, age, height, weight, function (pilot or cabin crew), habitual sleep duration, home base, typical commute time at home base, and diurnal type/chronotype (eveningness or morningness).

Procedures
Crew members collected data for 14 consecutive days. Both during normal flight duties and on days off, participants were asked to regularly fill in rating scales, to keep a sleep log, to continuously wear an actigraph (although not all participants were asked to wear one) and to perform reaction time tests (the latter only applied for pilots). Overall, this took them about 15 to 20 minutes per day.

During a typical flight duty day, participating crew members were asked to do the following:

After waking up:

- fill in the sleep log on the data collection app using their iPhone or iPad;
- provide KSS and SP ratings via the app; and
- complete the PVT via the app (pilots only).

At the start of their FDP:

- provide KSS and SP ratings 15 minutes prior to TOD for each sector. For long-haul flights[21], also rate KSS and SP during the cruise phase;
- complete the PVT 15 minutes prior to TOD of the final sector of the day (pilots only);
- press the button on the actigraph at each attempt to take a nap; and
- after taking a nap, fill in the sleep log via the app.

At the end of their FDP:

- rate mental effort using the slide bar;
- fill in the FDP worked;
- fill in hassle factors;
- press the button on the actigraph when going to sleep; and
- fill in the sleep log.

The 14 days of data collection started with two days off. The first day was used for familiarisation with the data collection app, while the second for gathering baseline measures. During all off-duty periods, besides wearing the actigraph (if they were given one), crew members were asked to provide KSS and SP ratings and to perform the PVT in the morning, in the afternoon and in the evening.

Crew members were offered training material via a dedicated website, to familiarise and train themselves with the use of the protocol and the app for data collection. They could contact the investigators via telephone or email in case of ambiguities. In addition, the airline coordinators were informed about the data collection details when questions were addressed to them.

[21] For this study long-haul flights were defined as flights longer than 5 hours.

Data handling

The participant data that was gathered was handled confidentially and anonymously, in compliance with the European General Data Protection Regulation.

The participant data that was collected through the application (tablet or smartphone) was transferred securely to the repository over an encrypted connection. The data (coded/de-identified) was stored in a password-protected central database repository that could only be accessed by project members.

Chapter 6: Analysis of field data

This chapter presents the analysis of the collected field data.[22]

Crew member data

The following 24 airlines accepted our invitation to participate in the field study:

Adria Airways, Air Baltic, Air Europa, Alitalia, ASL Airlines Belgium, BRA Braathens, British Airways, Cargolux, Condor, Croatia Airlines, Czech Airlines, Flybe, Iberia, KLM, Lufthansa, Lufthansa Cargo, Norwegian Air Int., Norwegian Air Shuttle, Scandinavian Airlines, SunExpress Deutschland[23], TAP Portugal, TAROM, Vueling, and WIZZ Air.

As to the geographical distribution of the airlines, six were from Eastern Europe, nine from western Europe, four from northern Europe, and five were from southern Europe. The distribution of the field study participants is unknown, due to the required anonymity of the airlines and aircrew.

Volunteers could register to participate in the collection of field data via an online portal. After registering, giving consent and familiarising themselves with the data collection app and measurement protocol, the volunteers started their 14 days of data collecting.

Figure 3 illustrates the flow from initial registrations, followed by dropouts, cancellations and insufficient quality of the collected data, and the resulting number of crew members proving adequate quality data. Data from a total of 2 877 FDPs were gathered. Figure 4 details the number of gathered FDPs per month during the data collection period, which extended from July 2017 to February 2018.

The crew members needed to be acclimatised[24] for the FDPs to be included in the data set. A crew member was considered acclimatised to a two-hour-wide time zone surrounding the local time at the point of departure. If the aircrew was not acclimatised during any part of the FDP, the corresponding duty was excluded from the data set. This resulted in 173 excluded FDPs.

Four further FDPs were excluded because they were performed with an augmented crew.

[22] This is an extract of D2.3 (Performance of the Data Collection and Data Analysis).
[23] SunExpress Deutschland volunteered participation without being invited explicitly.
[24] ORO.FTL.205 defines acclimatisation as a state in which a crew member's circadian biological clock is synchronised to the time zone where the crew member is.

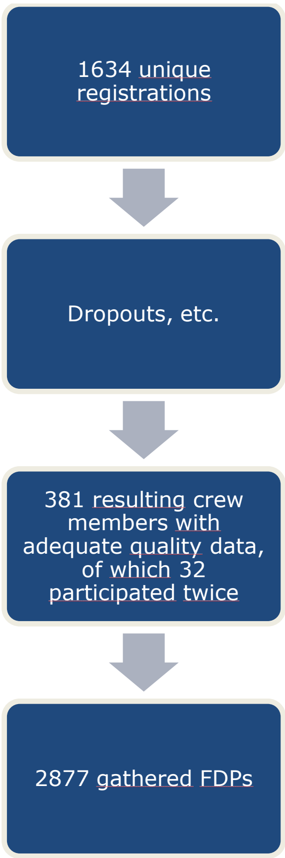

Figure 3 *Flow from registration to gathered FDPs. Out of 381 participants, 32 volunteered to take part in the 14 days of data collection twice (in two separate periods, not in one stretch).*

Table 1 and Table **2** provide information on the crew member demographics and type of operators that collected data.

Table 1 Crew member demographics

	Participants No (%)	**Age** Mean	**Gender** No (%)	**Habitual sleep length** Mean	**Home base commute time** Mean
Pilot	261 (68%)	40.9 yr	M 251 (91%) F 24 (9%) U 2 (0%)	469.3 min	121.9 min
Cabin crew	120 (32%)	37.0 yr	M 60 (44%) F 73 (54%) U 3 (2%)	473.8 min	113.6 min
Total	**381 (100%)**	**39.9 yr** **(SD 9.0)**	**M 311 (75%)** **F 97 (24%)** **U 5 (1%)**	**470.8 min** **(SD 46.8)**	**119.2 min** **(SD 59.6)**

Abbreviations: M: Male; F: Female; U: Unknown; SD: Standard Deviation; Min: Minutes; Yr: Years of age.

Table 2 Type of operator, as recorded by participants using the data collection app

	Participants No (%)	Pilot/Cabin crew
Network operator	173 (45%)	107/66
Point-to-point operator	130 (34%)	78/52
Cargo operator	78 (21%)	76/2
Total	**381 (100%)**	**261/120**

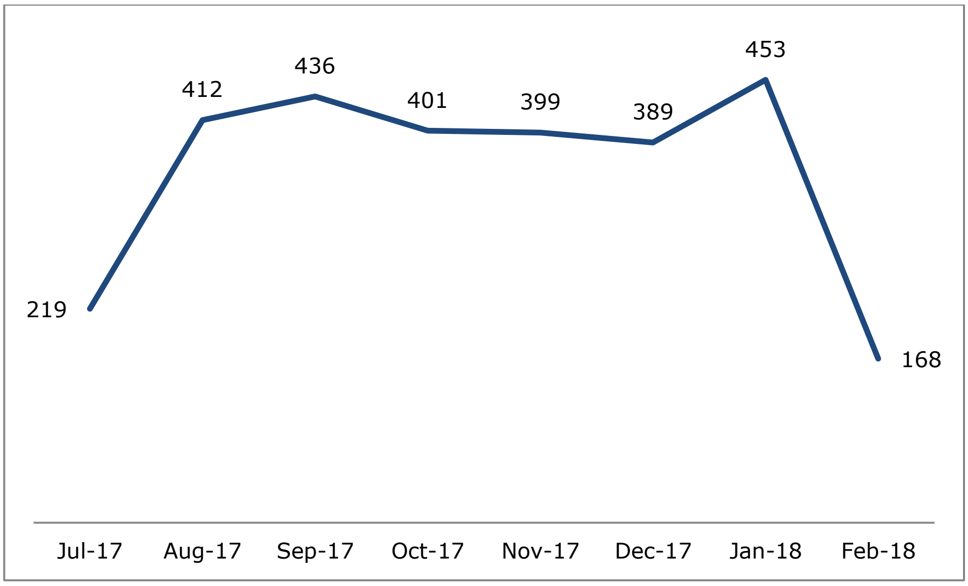

Figure 4 *No of gathered FDPs per month (2017-2018)*

The following tables present the gathered sample sizes for night FDPs longer than 10 hours, disruptive schedule FDPs, and the baseline data set.

Table 3 Sample sizes for FDP Night > 10 hours, FDP Disruptive schedules, and FDP Baseline set

	FDP Night duties > 10h No	**Full FDP Disruptive schedules set** No	**FDP Baseline** No
Pilot FDPs	136	822	1932
Cabin FDPs	65	354	945
Total FDP sample size	**201**	**1176**	**2877**

Table 4 Sample sizes for disruptive schedule FDPs: early starts, late finishes, and nights

	Early starts No	**Late finishes** No	**Nights** No
Pilot FDPs	181	147	494
Cabin FDPs	96	75	183
Total FDP sample size	**277**	**222**	**677**

Table 5 Sample sizes for consecutive disruptive schedule FDPs (at least two FDPs in a row): consecutive early starts, late finishes, nights, and mix

	Consecutive early starts No	**Consecutive late finishes** No	**Consecutive nights** No	**Consecutive mixes** No
Pilot FDPs	49	11	94	87
Cabin FDPs	31	12	29	26
Total FDP sample size	**80**	**23**	**123**	**113**

Table 6 Sample sizes for consecutive disruptive schedule FDPs (two or more FDPs in a row): consecutive early starts, late finishes, and nights

	Consecutive early starts No	**Consecutive late finishes** No	**Consecutive nights** No
2 in a row	56	18	90
3 in a row	18	4	21
4 in a row	5	1	8
5 in a row	1	-	4
Total FDP sample size	**80**	**23**	**123**

Table 7 Sample size disruptive schedule FDPs: Mix

	Early start - late finish No	**Late finish - night** No	**Night - early start** No	**Late finish - early start** No	**Night - late finish** No	**Early start - night** No
Total FDP sample size	**11**	**19**	**18**	**2**	**37**	**26**

Crew data representativeness and sample size

The 24 airlines participating in the data collection were used as a reference set for the EU aviation sector as a whole. Our examination of these participants initially suggested an overrepresentation of the eastern region in the data set, with six airlines participating. However, the six airlines are small relative to the others, which is why, also based on the geographical distribution and type of operations included, we consider this representation a fair one. It is thus appropriate to use the set as a proxy for the EU aviation sector. This conclusion was confirmed by our estimates of the size and geographical distribution of the entire EU aircrew population.

Data was collected on 2 877 FDPs by 381 crew members.[25] The participating population consisted of 68 % pilots and 32 % cabin crew, whereas in the entire EU crew population approximately 59 % are cabin crew. The relatively high share of pilot participation was partly due to the inclusion of three cargo operators. This was by design, as we sought airlines with regular night FDPs and disruptive FDPs. The cargo operators only had a small number of cabin crew (2 members), presumably because there is just a small number of cabin crew employed by the operators.

Data was collected for a period of eight months. During this period (from July 2017 through February 2018) both low- and high-workload periods for the airlines were covered.

[25] That is approximately 0.3 % of the entire crew population base in Europe as estimated in D2.2 (Definition of the Data Collection Process) in chapter 2, page 11.

Data analysis: identifying fatigue hotspots

In this part of the research, the primary analyses were performed using KSS values, scored at TOD during the final sector of the FDP under consideration. Again, data analysis consisted of three steps: checking for high fatigue scores, comparing fatigue scores between FDP categories, and finding clusters of variables.

Step 1: Check for high fatigue scores

We sought to identify whether high fatigue scores occurred in night FDPs longer than 10 hours and in disruptive schedule FDPs. A high level of fatigue was defined by a KSS score ≥ 7 and an SP score ≥ 6. Total sleep in the 24 hours prior to TOD (Sleep24h) and napping during FDP (FDPsleep) for the high (KSS ≥ 7) and low (KSS < 7) levels of fatigue were examined. The percentages of high and low fatigue scores (KSS) for each hour of the day were also assessed.

Step 2: Compare fatigue scores between FDP categories

Differences were calculated between the KSS scores in night FDPs longer than 10 hours and those in control FDP categories, and between disruptive schedule FDPs and control FDP categories.

The primary approach for analysing night FDPs longer than 10 hours was to compare short FDPs (≤ 10h) with long (> 10h) ones with respect to the level of fatigue at TOD, adjusting for factors that may influence the outcome. The same analysis was repeated using, instead of a cut-off of 10 hours, cut-offs of 8, 9, 11 and 12 hours.

For disruptive schedule FDPs, all disruptive FDPs were compared with all non-disruptive (essentially daytime) FDPs. Additional comparisons were performed for the different types of disruptive schedule FDPs (i.e. early starts, late finishes, and nights) and between one disruptive FDP and two consecutive FDPs.

Step 3: Find clusters of variables

We used multiple logistic regression models to determine clusters of FDP-related characteristics (predictors) under which high levels of fatigue occurred, also referred to as 'fatigue hotspots'.

Main results

Fatigue at TOD during the FDPs of interest

Figure 5 illustrates point estimates of the probability of occurrence of high fatigue at TOD (KSS ≥ 7) during the FDPs of interest and in the entire data set (i.e. all FDPs collected, referred to as the baseline). Compared with the baseline, the point estimates clearly increased for long and short night FDPs (duration > 10h and ≤ 10h), for night FDPs (including all night FDPs), and for late finishes FDPs. A marginal increase was found for mixed FDPs. The mixes represented the following combinations: an early start FDP preceded by a late finish FDP; an early start FDP preceded by a night FDP; a late finish FDP preceded by an early FDP; a late finish FDP preceded by a night FDP; a night FDP preceded by an early start FDP; and a night FDP preceded by a late finish FDP. No increase was found for early start FDPs. It must be borne in mind that these results are only descriptive, i.e. they do not represent a statistical comparison between the FDPs of interest and the baseline. Also, the baseline represents the mean level across all FDPs (not, for example, daytime FDPs only).

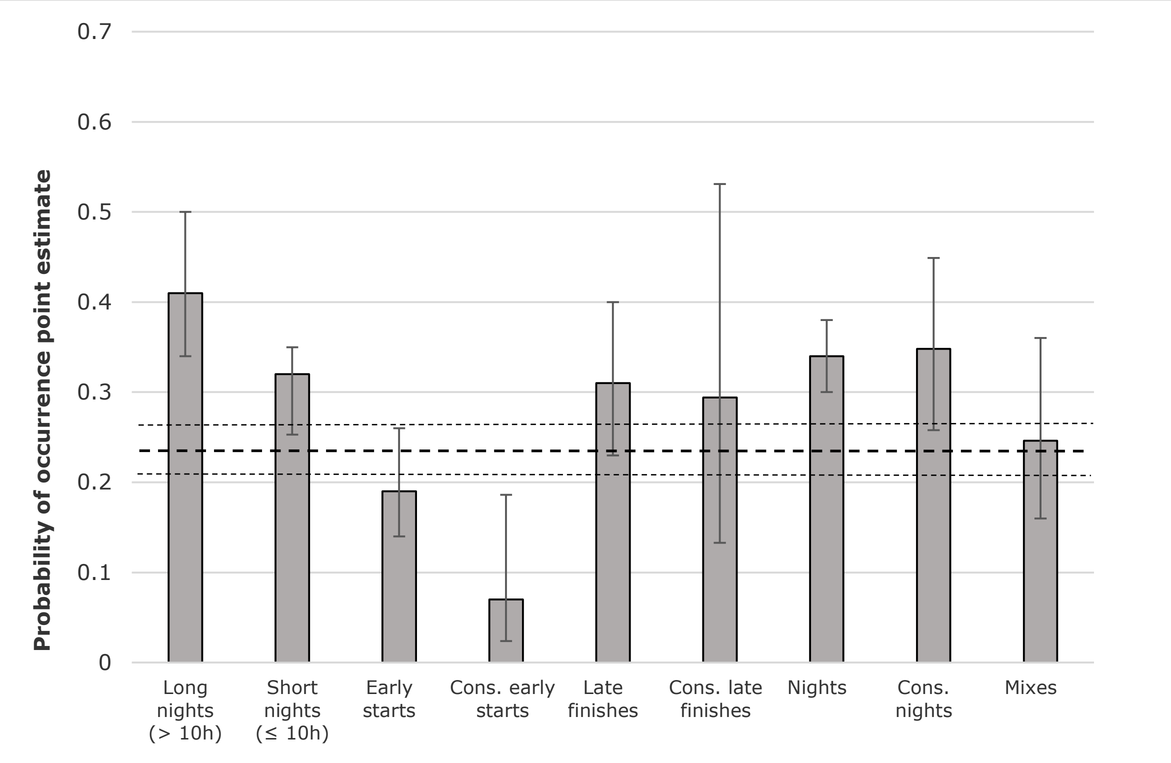

Figure 5 *Point estimates for the probability of occurrence of high fatigue at TOD during the FDPs of interest and the baseline condition (all FDPs collected and denoted by the thick dashed horizontal line). The thin dashed lines denote the 95 % confidence interval (CI). The vertical lines indicate 95 % CIs of the FDPs of interest. The number of observations by FDP type are as follows: long nights (> 10h) 146, short nights (≤ 10h) 348, early starts 163, consecutive early starts 43, late finishes 123, consecutive late finishes 17, nights 494, consecutive nights 92, and mixed combinations of disruptive schedules 69.*

Figure 6 shows odds ratios (ORs) for high fatigue (KSS ≥ 7) at TOD during the FDPs of interest. These analyses were based on between-subjects data extracted from the entire data set (baseline): the reference used was all daytime FDPs (all FDPs with start time ≥ 7.00 and end time < 23.00).

The OR for high fatigue during long-night FDPs (duration > 10h) was not higher compared with short-night FDPs (duration ≤ 10h). The comparisons between disruptive-type FDPs (early start, late finish, and night FDPs) and daytime FDPs yielded significantly higher ORs. This was especially true for late finish and night FDPs, and less so for early start FDPs.

It must be borne in mind that the data presented in Figure 6 does not allow for a fair comparisons to be made between the FDPs of interest, but only between the FDP of interest and the reference condition. Also, the reference condition for long-night FDPs is short-night FDPs, unlike for the other FDPs of interest, which use daytime FDPs as their reference.

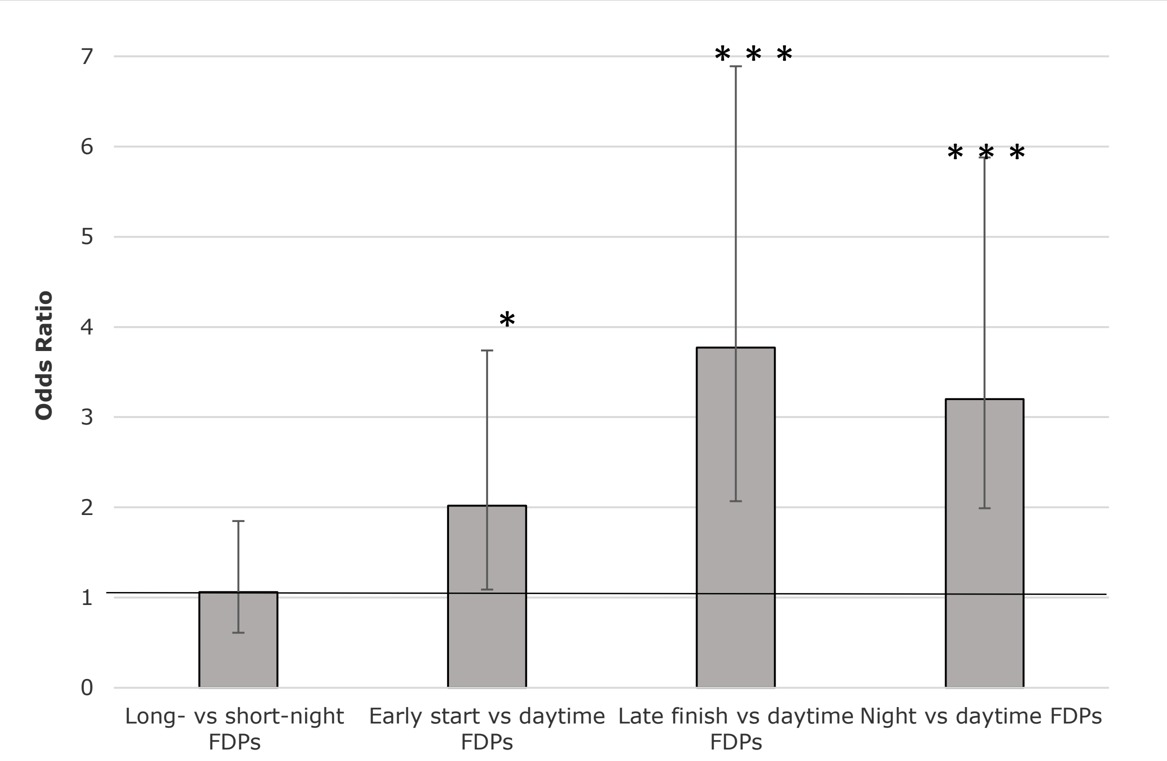

Figure 6 *ORs for reporting high fatigue at TOD (KSS ≥ 7) during the FDPs of interest compared with their reference conditions. The horizontal line denotes the reference FDP category. Please note that the reference condition for long-night FDPs is short-night FDPs, unlike for the other FDPs of interest, which use daytime FDPs as their reference. The vertical lines indicate the 95 % CIs. A value greater than 1 indicates an increased OR. * = p < 0.05; *** = p < .001.*

Figure 7 shows the results of a supplementary analysis that compared long- and short-night FDPs in more detail using a between-subjects data set extracted from the baseline. No significant differences in fatigue at TOD were observed for long-night FDPs when the criterion for long duration was incrementally increased from > 8 hours to > 12 hours.

Figure 8 shows the mean KSS values at TOD during the first and second consecutive FDPs of interest. None of the comparisons within each FDP type (early start, late finish, night) suggests an increase in fatigue from the first to the second FDP in a row.

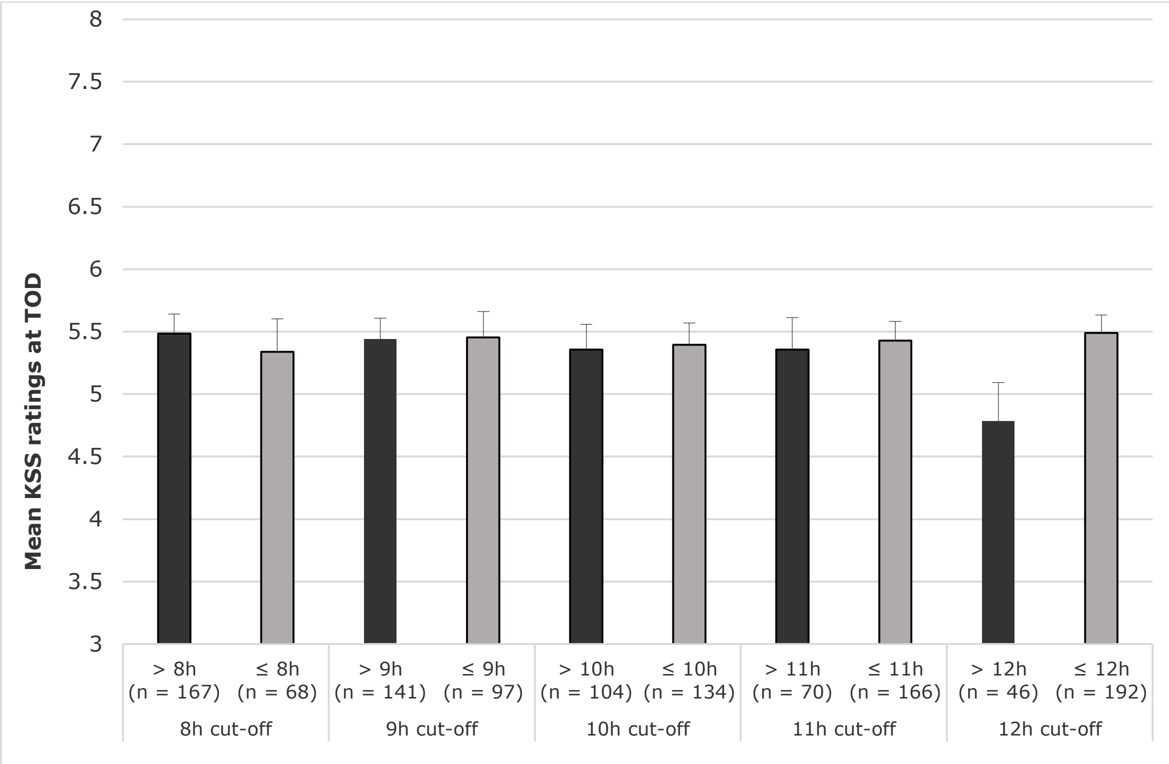

Figure 7 *Mean KSS ratings at TOD for the long- and short-night FDPs. The criterion for the duration of a long-night FDP ranged from > 8h to > 12h. The black bars represent long FDPs and the grey bars short FDPs in the between-subjects data. The vertical lines denote the standard errors. Note that the y-axis covers only part of the 9-point KSS.*

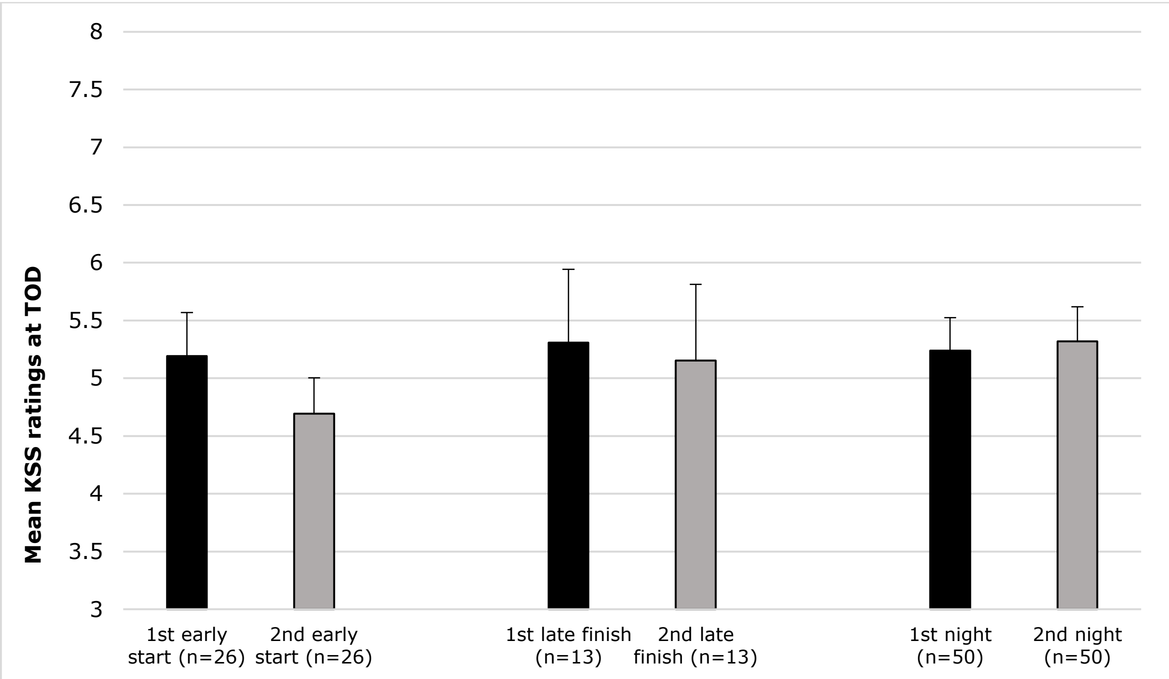

Figure 8 *Mean KSS ratings at TOD for the first (black bars) and second (grey bars) consecutive early start (n = 26), late finish (n = 13), and night (n = 50) FDPs in the within-subject data. The vertical lines denote the standard errors.* Please note that the *y-axis covers only part of the 9-point KSS.*

Predicting high fatigue at TOD during the FDPs of interest

Figure 9 shows the main FDP-related predictors of high fatigue at TOD. The results are based on data sets that include an FDP of interest and its reference FDP (daytime/non-disruptive FDPs). The results of the multiple regression analysis indicate that the increased odds of high fatigue at TOD during early start FDPs are attributable only to the earlier start time itself. When early start FDPs were analysed without their reference condition, none of the FDP-related characteristics (including prior sleep) explained the occurrence of high fatigue at TOD.

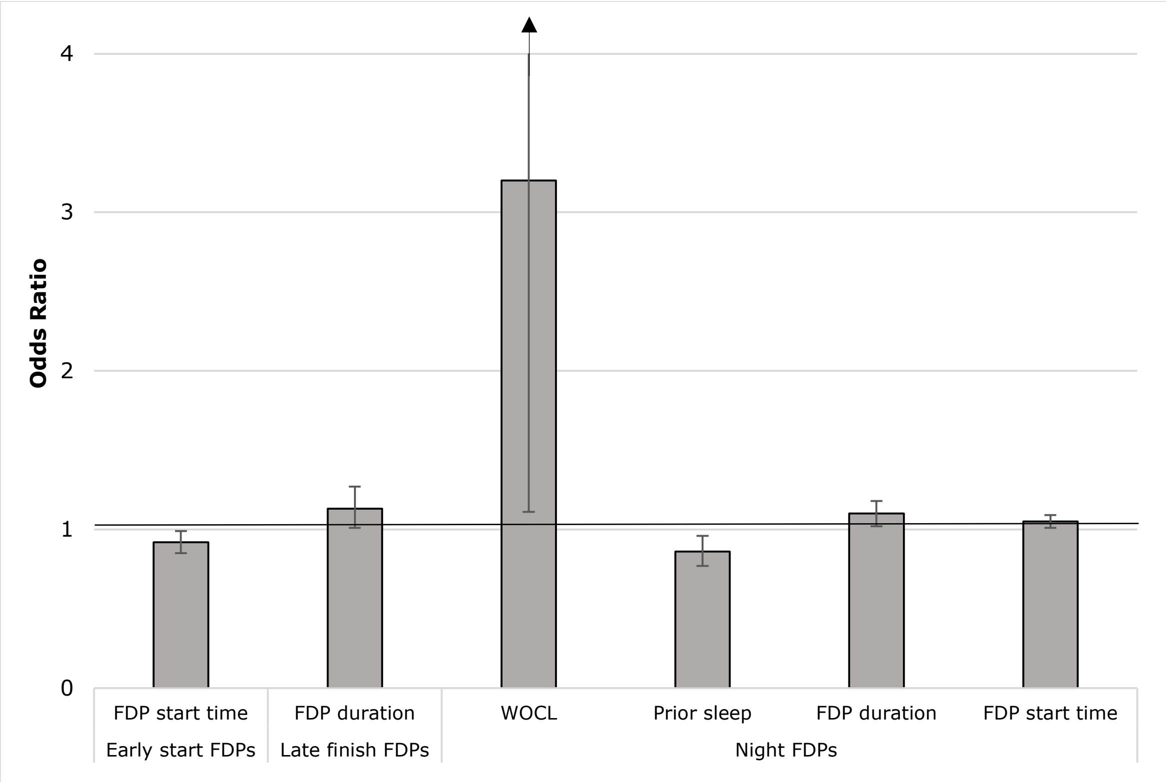

Figure 9 *ORs of high fatigue at TOD (KSS ≥ 7) for the FDPs of interest and the reference FDP (daytime). The horizontal line denotes the reference FDP (all FDPs with start time ≥ 7.00 and end time < 23.00). The vertical lines indicate the 95 % CI. A value greater than 1 indicates an increased OR. Note that the upper limit of the 95 % CI for the WOCL (window of circadian low) falls outside the y-axis scale (9.19), which is indicated by an arrow.*

The increased odds of high fatigue at TOD for late finish FDPs is to some extent attributable to the longer FDP duration and later FDP finish time. A supplementary simple regression analysis without the reference FDP found that FDP start time (OR = 0.82 (CI = 0.71; 0.94), $p = .004$), FDP duration (OR = 1.20 (CI = 1.04; 1.39), $p = .012$), and FDP end time (OR = 0.54 (CI = 0.31; 0.95), $p = .033$) were significantly associated with the occurrence of high fatigue at TOD.

For night FDPs, the main predictors of increased odds of high fatigue at TOD were 'on duty during WOCL' (2.00 – 5.59) and 'shorter prior sleep in the previous 24 hours'. Longer sleep acted as a protective factor, as indicated by an OR of < 1 (see Figure 9). After removing these two factors from the regression model, longer FDP duration and later FDP start time became significant predictors. Besides these FDP-related characteristics, being a cabin crew member (versus being a pilot) was associated with increased odds of high fatigue at TOD for night FDPs (OR 1.76; 95 % CI 1.07 - 2.90, $p < 0.05$).

In a supplementary analysis with no reference FDPs, only the amount of sleep in the previous 24 hours and being a cabin crew member (versus being a pilot) significantly predicted the occurrence of high fatigue at TOD during night FDPs (prior sleep: OR 0.84; 95 % CI 0.75 - 0.94, $p < 0.01$; cabin crew member: OR 1.89; 95 % CI 1.07 - 3.36, $p < 0.05$). The effect of WOCL could not be analysed specifically because the WOCL was the basis of the definition of night FDPs.

Additional analyses with reference to FDP duration

As FDP duration showed very limited predictive power regarding high fatigue at TOD for night FDPs, we made an additional attempt to examine this factor in depth. First, we divided night FDPs into two categories using 6.00 as the FDP end time cut-off. A regression analysis performed separately for these two night FDP categories did not, however, show that FDP duration were associated with fatigue at TOD (end time < 6.00: OR = 1.04, CI = 0.86 - 1.27; end time ≥ 6.00: OR = 0.85, CI = 0.63 - 1.13).

Secondly, we looked at variation in duration of night FDPs. Only two turned out to be shorter than four hours. We did find an association between FDP duration and end time: night FDPs ending in the morning or forenoon were longer than FDPs ending at night. Both factors likely reduced the effect of FDP duration on fatigue in the present data.

Finally, we analysed the relationship between FDP duration and high fatigue at TOD for daytime FDPs (those starting and ending between 7.00 and 23.00). A simple regression analysis found only FDP duration (OR = 1.24, CI = 1.09 - 1.41, $p = 0.001$) and end time (OR = 1.13, CI = 1.00 - 1.28, $p = 0.045$) to be significant predictors. Entering the two predictors into a multiple logistic regression, we obtained a reduced, but still significant, result for FDP duration (OR = 1.21, CI = 1.04 - 1.39, $p = 0.012$). The range in FDP duration was considerable for daytime FDPs (0.75h - 13.8h), which is in contrast to the overall scarcity of short FDPs in the night FDPs. This contrast might contribute to the difference found between night FDPs and daytime FDPs in the association between FDP duration and high fatigue at TOD.

Additional results using alternative FDP categories

For the analyses described above, FDPs were classified based on the criteria described in ORO.FTL. Our results suggest that this classification may not be optimal. We therefore explored an alternative way of classifying FDPs based on their start and end times.

We made two changes to the classification of the FDPs of interest. First, we created two categories for early start FDPs: early start FDPs (start time between 5.00 and 6.59) and deep early start FDPs (start time between 2.00 and 4.59). This division enables us to determine if starting time is a significant factor in early start FDPs. In ORO.FTL, all FDPs starting between 2.00 and 4.59 are considered night FDPs, though the crew is usually able to obtain at least some night sleep just prior to their FDP.

Second, we created two categories for night FDPs: night FDPs (ending between 2.00 and 5.59) and deep night[26] FDPs (start time 1.59 or earlier, end time 6.00 or later). This division enables us to determine if night FDPs that end within the WOCL differ

[26] Alternative ways of addressing these deep nights might be 'late nights', 'full night', 'WOCL night', or 'nights that encompass the WOCL'.

from those that end after the WOCL (and thus completely cover the WOCL). This new classification is closely linked to the three-process model of alertness, which is a scientifically established model to predict sleep and fatigue (Åkerstedt & Folkard, 1997).

Figure 10 shows point estimates for the probability of occurrence of high fatigue at TOD in each of the FDP categories, compared with non-disruptive FDPs (starting 7.00 or later and ending at 22.59 or earlier). Each FDP of interest showed an increased tendency towards high fatigue. This tendency was especially discernible during late finish, night, and deep night FDPs. Among these, the tendency was most pronounced in the deep night category. The results for the early start and deep early start FDP categories were very similar.

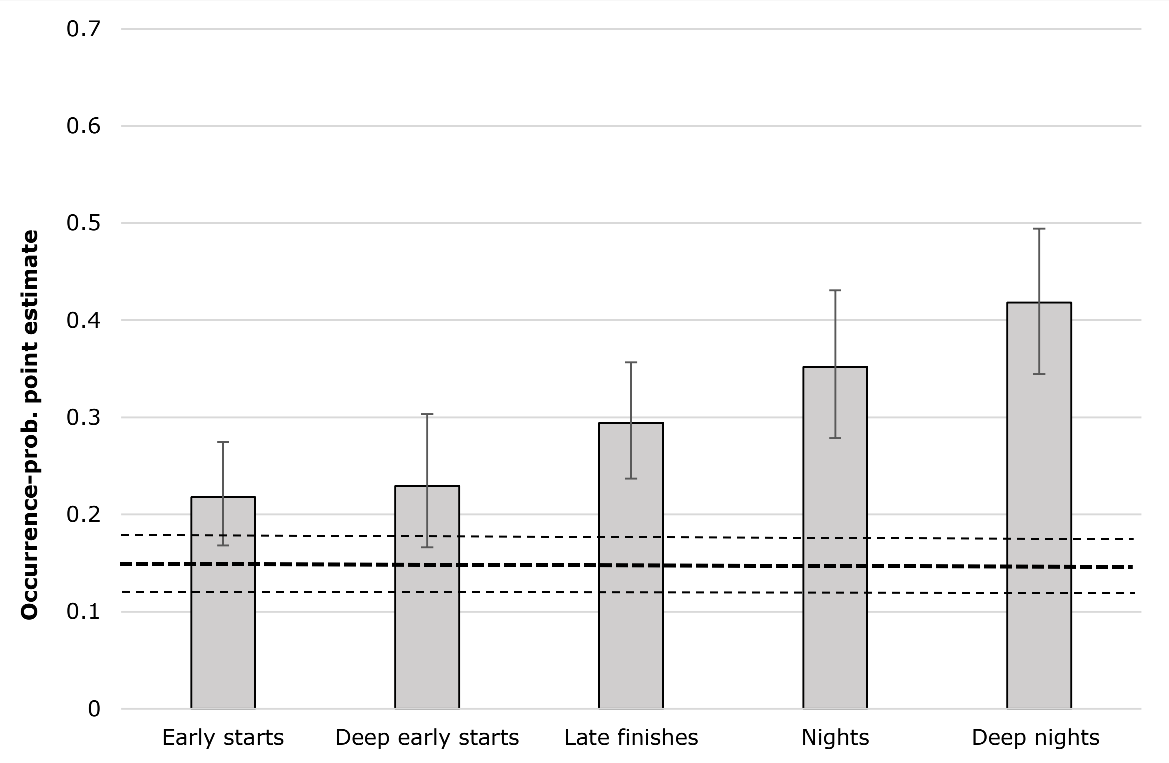

Figure 10 *Point estimates for the probability of occurrence of high fatigue at TOD in each FDP category of interest. The vertical lines denote a 95 % CI. The thick dashed horizontal line denotes the daytime FDP category (all FDPs with start time ≥ 7.00 and end time ≤ 22.59). The thin dashed lines represent the 95 % CI.*

Why are late finish and night FDPs particularly fatiguing? One tentative explanation is a reduced sleep-wake ratio. In the FDPs of interest, that ratio fell clearly below the level of the daytime FDPs (mean 0.57; 7.35 hours of sleep followed by 12.85 hours of wakefulness) especially in the two night FDP categories (mean 0.30; 0.22). It also bears mentioning that the deep night FDPs were exceptionally long in duration (mean 10.10 hours), which may also explain the result. A tentative reason for not finding a difference between early start and deep early start FDPs was the unexpectedly favourable sleep-wake ratio in both categories (mean 0.50).

Table 8 shows logistic regression results for the different FDP categories. In each of these categories, the OR of high fatigue at TOD was increased compared with daytime FDPs. The most pronounced increase was found in the deep night category. In this category, the OR was about two times higher than in the other four categories.

Table 8 Logistic regression results predicting high fatigue at TOD in the alternative FDP categories. The analyses are based on between-subjects data. Daytime FDPs serve as the reference condition.

FDP category	***OR***	***CI***	***p***	***N*** *daytime/disruptive*
Night	4.16	2.00;8.65	0.000	165/51
Deep night	8.04	3.58;180	0.000	154/63
Early	3.28	1.30;8.25	0.012	174/39
Deep early start	4.16	1.63;10.22	0.000	170/30
Late finish	4.65	2.08;10.40	0.000	190/53

OR = Odds Ratio; CI = Confidence Interval; p = level of significance. High fatigue was defined by scores on the KSS equal or higher than 7.

Summary of the results

These results are based exclusively on the current field data and are discussed in detail in the section 'Discussion and conclusions'. The point estimates for the probability of occurrence of high fatigue at TOD during the FDPs of interest and the baseline condition as referred to in the summary tables below are presented in Figure 5 and Figure 10.

*Summary of results on high fatigue at TOD for **non-consecutive** FDPs*

FDP of interest	***Main results on high fatigue[1] at TOD and its predictors***
Night duties (> 10h)	All night FDPs were associated with high probability of high fatigue at TOD. This probability was similar for short (≤ 10h) and long nights (> 10h).
Nights	The probability of high fatigue at TOD during night FDPs was higher than during daytime FDPs. Encroachment of the FDP on the WOCL and shorter prior sleep were the significant predictors. When night FDPs were analysed alone, shorter prior sleep explained the occurrence of high fatigue at TOD.
	To cover the continuum from evening to night, late finish plus night (start time before 24.00) FDPs were combined. A higher probability of high fatigue at TOD during these FDPs than during daytime FDPs was predicted by encroachment on the WOCL, shorter prior sleep, later FDP start time, and longer FDP duration. When late finish plus night FDPs were analysed alone, encroachment on the WOCL, earlier FDP end time, and shorter prior sleep explained the occurrence of high fatigue at TOD.
	To cover the continuum from late night to early morning, very early (3.00 – 4.59) and early (5.00 – 06.59) starting FDPs were combined[2]. A higher probability of high fatigue at TOD during these FDPs than during daytime FDPs was explained by earlier FDP start time and shorter prior sleep. When these FDPs were analysed alone, only shorter prior sleep explained the occurrence of high fatigue at TOD.
	An alternative way of classifying FDPs was suggested. When applying this classification, the probability of high fatigue at TOD was found to be similar for deep early (start time 2.00 – 4.59) and early (start time 5.00 – 6.59) start FDPs[3]. The highest probability of high fatigue at TOD was found for deep night FDPs that covered the entire night (start time 1.59 or earlier, end time 6.00 or later).
Early starts	The probability of high fatigue at TOD during early start FDPs was higher than during daytime FDPs. Earlier FDP start time was the only statistically significant predictor. When early start FDPs were analysed alone, none of the FDP-related characteristics explained the occurrence of high fatigue at TOD.
Late finishes	The probability of high fatigue at TOD was higher during late finish FDPs than during daytime FDPs. A longer FDP duration was the only significant predictor. When late finish FDPs were analysed alone, a longer FDP duration, an earlier FDP start time, and an earlier FDP end time explained the occurrence of high fatigue at TOD.

*Summary of results on high fatigue at TOD for **two consecutive** FDPs*

FDP of interest	***Main results on high fatigue[1] at TOD***
Consecutive early starts	Fatigue levels at TOD were similar for the first and second consecutive early start FDPs.
Consecutive late finishes	It seemed that the probability of high fatigue at TOD was similar for the first and second consecutive late start FDPs.
Consecutive nights	Fatigue levels at TOD were similar for the first and second consecutive night FDPs.
Mix	It seemed that the probability of high fatigue at TOD during mixes of disruptive schedule FDPs was greater than the corresponding probability in the baseline data set, containing all FDPs collected.

[1] A high level of fatigue was defined as KSS scores equal to or greater than 7 (= sleepy, but no effort to keep awake).
[2] FDPs starting between 3.00 and 4.59 are considered night FDPs in the current FTL, whereas those starting between 5.00 and 06.59 are not.
[3] The deep early FDPs are considered night FDPs in the current FTL, whereas early FDPs are not.

Assessment of the strength of the evidence found

We assessed the strength of the evidence found mainly based on the degree of alignment (determined by the researchers) between the results of the field study and those of the roster modelling, online survey, and relevant prior research (benchmarking exercise[27]). Also, the data sample sizes for the particular FDP types were considered. Table 9 summarises these results.

Table 9 Strength of the evidence found

FDP of interest	***Strength of the evidence collected***
Night duties (> 10h)	***Strong***: i) strong alignment with roster data, survey and benchmarks; ii) sufficient sample sizes
Nights	***Strong***: i) strong alignment with roster data, survey and benchmarks; ii) sufficient sample sizes
Early starts	***Moderate***: i) moderate alignment with roster data, survey and benchmarks; ii) sufficient sample sizes
Late finishes	***Strong***: i) strong alignment with roster data, survey and benchmarks; ii) sufficient sample sizes
Consecutive early starts	***Moderate***: i) moderate alignment with roster data, survey and benchmarks; ii) insufficient field data sample size
Consecutive late finishes	***Weak***: i) inconsistent alignment with roster data, moderate alignment with survey, and strong alignment with benchmarks; ii) insufficient field data sample size
Consecutive nights	***Strong***: i) strong alignment with roster data, survey and benchmarks; ii) large sample sizes
Mix	***Weak***: i) moderate alignment with roster data, survey and benchmarks; ii) sufficient sample sizes but insufficient field data on *specific* mixed combinations of disruptive schedules

The assessment of the strength of the evidence provides no reason to rebuke any of the main results as presented above. The main results already accounted for the weak and moderate levels of the evidence provided.

Discussion and conclusions

Understanding the main analyses results

Our field study showed that the probability of high levels of fatigue at TOD is high during night and late finish FDPs, among both pilots and cabin crew. For early start FDPs and mixed combinations of disruptive schedules, our findings were less clear.

It is important to note that our results are based on crew fatigue ratings at the TOD of the final sector of an FDP. To overcome this limitation, we also conducted some additional analyses that considered the highest fatigue rating crew made during either the cruise phase or TOD at any sector[28]; i.e. not just in the final sector. These results appeared to be well aligned with those that utilised only the ratings given at TOD of the final sector.

No significant difference in fatigue at TOD was found between night duties longer than 10 hours, compared with shorter night FDPs. This result does not, however, mean that FDP duration is not an important determinant of fatigue. The main reason for the result probably is that high fatigue during night FDPs is mainly caused by the unfavourable time of the day (circadian factor) and a reduced sleep-wake ratio

[27] Presented in D3.2 (Benchmark of this Analysis with Other Relevant Resources).
[28] Presented in D2.3 (Performance of the Data Collection and Data Analysis) in chapter 3, page 18.

(homeostatic factor). These two factors likely interacted with the influences of FDP duration. In addition, night FDPs seldom are of short duration (i.e. in the field data set, 1.5 % of the night FDPs were found in the ≤ 4h category; in the roster data set this was 8.6 %), which limits the range of variation of this FDP characteristic. It is also important to note that we did not measure the length of time participants were fatigued. This limitation can be assumed to underestimate FDP duration as an underlying factor for fatigue in our analyses.

No significant difference in fatigue at TOD was found between the first and second consecutive disruptive FDPs. Unfortunately, our field data did not permit us to study cumulative fatigue over sequences longer than two consecutive FDPs. This might be the result of the current regulatory fatigue management controls and/or company rostering rules. The roster data[29] also showed relatively low sample sizes for the different types of consecutive disruptive schedules, which is especially the case for four or more disruptive schedules in a row. This is likely associated with the required extension of the recovery rest period if a crew member performs four or more disruptive schedules (CS.FTL.1.235 Rest Periods). The same lack of data holds for schedules where an early start FDP is preceded by a duty sequence that compromises sleep (e.g. quick transitions). This limitation restricts our possibilities to explore the fatigue associated with the different types of disruptive schedules.

The strongest predictors of higher probability of high fatigue at TOD, as compared with daytime FDPs, varied by FDP type. For early start FDPs, only an earlier start time was a significant predictor. For late finish FDPs, the only significant predictor was a longer FDP duration. In the case of night FDPs, the pertinent predictors were encroachment on the WOCL (2.00 – 5.59), short prior sleep, and being a cabin crew member.

The difference between pilots and cabin crew is of interest. This could be explained by the possibility that cabin crew may have been more fatigued. Another explanation for the result lies in a difference in the level of workload at TOD between the two crews. At TOD, cabin crew are typically sitting in the cabin crew jump seat after a potentially busy work period in the cabin. In contrast, at TOD pilots are in a high workload phase of the flight, having just finished preparing for descent, approach and landing, and commencing descent. Unlike function, age was not a significant individual factor, as it showed some predictive power only during daytime FDPs. The other individual factors examined – diurnal type, habitual sleep length, body mass index, and commuting time – were not significant predictors of high fatigue during any FDP type.

Interestingly, the FDP-related characteristics were rather weak predictors of high fatigue at TOD (KSS ≥ 7) for the early start, late finish and night FDPs when each FDP category of interest was analysed alone (i.e. without combining it with daytime FDPs). This finding suggests that a simple FDP limit based on a characteristic such as FDP start time may not effectively control the likelihood of high fatigue at TOD, provided that the adjustment occurs within the limits set for that characteristic in the current analysis.

Implications for fatigue mitigation

Night duties longer than 10 hours

Our results suggest that increased fatigue at TOD during night duties longer than 10 hours may be difficult to effectively control by merely adjusting the FDP duration, since there are multiple other more influential determinants. There are other non-

[29] Presented in D2.1 (Identification of Potential Fatigue Hotspots) in chapter 2, page 18.

schedule related strategies for reducing fatigue at TOD during (long) night FDPs[30]. One of them is strategic sleep before and during a flight. For example, in-flight sleep during long night flights with augmented flight crew has been found to be beneficial to fatigue at TOD[31].

The use of rest before or during FDPs is supported by our findings about the frequency of napping on the flight deck (none of the flights was operated with an augmented crew and it was not recorded whether the napping was done under a controlled rest procedure). The frequency was higher than expected, as napping on the flight deck is only allowed to manage unexpected fatigue and to reduce the risk of fatigue during higher workload periods later in the flight. This behaviour was frequent especially during night flights longer than 10 hours (27 %). This kind of napping is not a substitute for proactive fatigue management via scheduling, sufficient pre-duty sleep or augmentation to enable sleep opportunities during a flight. Napping on the flight deck (under a controlled rest procedure) is currently considered as a reactive strategy to mitigate unexpected fatigue experienced during a flight.

Finally, it is worth remembering that these suggestions to mitigate fatigue during night duties of more than 10 hours focus solely on fatigue at TOD. We did not measure the length of time a crew member was fatigued during night FDPs. In other words, the duration of exposure to fatigue hazard remained unclear in the present study.

Early starts

Of all the disruptive duties, early starts turned out to be associated with the lowest fatigue scores at TOD. The analyses based on the entire data set did not show significant findings, whereas high fatigue on early starts was twice as likely as during daytime FDPs in the between-subjects data. We also attempted to include a part of the night FDPs, called 'deep early starts' (start time between 2.00 and 4.59), in the early starting FDPs. This, however, did not yield results that would have markedly differed from the original ones. The fact that the evidence of increased fatigue at TOD was uncompelling can be explained by two factors: i) a relatively good ratio between prior sleep and wake, largely due to the crew having been awake for a relatively short time at TOD at the end of an early duty; and ii) the final TOD does not encroach on the WOCL. In addition, our data did not cover the first two phases of a flight (blocks-off and top of climb). It is possible that fatigue is actually higher during this part of an early start FDP because of the influence of the WOCL.

Late finishes

The evidence of high fatigue during late finish FDPs was quite solid. Based on our results, it is worth considering further measures to curb fatigue during late finish FDPs. In particular, the observation that the results of the late finish FDPs were very similar to those of the night FDPs supports this conclusion.

Nights

Our results demonstrate the need to further mitigate fatigue while flying during the night. When considering mitigation strategies, it is important to note that the present study did not reveal FDP-related characteristics (except for encroachment on the WOCL) that predicted high fatigue at TOD during night FDPs. In other words, fatigue at TOD was independent of the FDP characteristics (e.g. start and end times, duration), as long as an FDP fell into the 'night FDP' category. Given this result, it is

[30] Wesensten et al. 2015; Gander, 2015; Dawson & McCulloch, 2005
[31] Van den Berg et al. 2016; Gander et al. 2013

difficult to propose any scheduling-based solution to mitigate high fatigue at TOD during night FDPs.

Our additional analyses revealed that especially deep night FDPs (end time after the WOCL and start time before the WOCL) involved high fatigue at TOD. This finding suggests that deep night FDPs need special attention when mitigating duty fatigue.

When interpreting our results of night FDPs, it is important to notice that high fatigue is to some extent an inevitable part of night work across industries, because human beings are day-oriented.[32].

Need to revise regulations?

We presented an example of an alternative way to categorise FDPs typical of disruptive schedules. First, we re-categorised the night FDPs starting between 2.00 and 4.59 as ‘deep early starts’. A reason for this change was that crew were able to obtain some night sleep just before these very early start FDPs, unlike the other night FDPs. Another reason was to determine if deep early FDPs involved more fatigue than early FDPs (start time between 5.00 and 06.59).

This re-categorisation revealed no sizable difference in high fatigue at TOD between deep early starts and early starts. Perhaps the relative favourable sleep-wake ratio (0.5) prior to TOD played a role in reducing fatigue during the deep early start FDPs. It must be noted, however, that fatigue was not measured at the beginning of the FDPs (e.g. at blocks-off or top of climb). This might be of importance, since sleepiness in the circadian rhythm peaks at about the same time as deep early FDPs start.

We similarly divided night FDPs into ‘nights’ (end time within the WOCL) and ‘deep nights’ (end time after the WOCL and start time before the WOCL) and found an exceptionally high rate of fatigue during the latter. A tentative explanation lies in three observations: the sleep-wake ratio was very low (0.22), the FDPs encompassed the WOCL, and FDP duration was particularly long.

In summary, our results suggest that late finish and night FDPs are more fatiguing than early start FDPs. In addition, deep night FDPs seem to be more of a concern than late finish and night FDPs. Our view is that the current definitions of FDPs typical of disruptive schedules could be more closely aligned to an established and science-based model used to predict fatigue (e.g. the three-process model of alertness referenced earlier). This revision would probably pave the way to better fatigue management.

[32] Åkerstedt, 1988; Monk, 1990; Sallinen & Hublin, 2015

Chapter 7: Conclusions and recommendations

This chapter presents the conclusions and recommendations drawn from the first and current FTL research phase.[33]

Conclusions

Night FDPs, both longer and shorter than 10 hours, were associated with a high probability of high fatigue at TOD. This is not fully reflected in the current FTL regulation and guidance material. The regulation and guidelines explicitly note the need for appropriate fatigue risk management and the importance of obtaining sufficient sleep in relation to night duties *longer* than 10 hours, but not for those *shorter* than 10 hours.

Within the FDPs defined as 'night' FDPs in the current regulations, three subgroups can be distinguished based on the probability of occurrence of high fatigue at TOD:

1. FDPs starting between 2.00 and 4.59;
2. FDPs ending between 2.00 and 5.59 and starting at 1.59 or earlier; and
3. FDPs ending at 6.00 or later and starting at 1.59 or earlier.

The existence of these subgroups is not recognised in the current FTL. Distinguishing these subtypes could help operators to design effective fatigue risk management strategies.

The factors that best predicted increased odds of high fatigue at TOD varied by FDP type. This suggests that fatigue mitigation measures should be based on various fatigue management strategies and tailored to FDP type and operator context.

Recommendations

The existence, frequency, and characteristics of the FDPs of interest vary with each operator. It is therefore not possible to provide precise recommendations that would be of equal effectiveness for every operator. However, each operator may manage the fatigue risk of its specific characteristics of its schedule effectively by following the principles of fatigue risk management. Therefore, the recommendations are aimed at fatigue risk management.

The following recommendations are made in no particular order.

Recommendation 1

Within the FDPs that are defined as 'night' FDPs in the current regulation, three subgroups can be distinguished based on the probability of occurrence of high fatigue at TOD:

- FDPs starting between 2.00 and 4.59;
- FDPs ending between 2.00 and 5.59 and a start at 1.59 or earlier; and
- FDPs ending at 6.00 or later and starting at 1.59 or earlier.

It is recommended that these subtypes be included in the definition of night FDPs to help operators to design effective fatigue risk management strategies.

[33] This is an extract of D3.1 (Analysis of the Fitness for Purpose of the Current Safety Management Controls), D3.2 (Benchmark of this Analysis with Other Relevant Sources), and D3.3 (Conclusions and Recommendations).

Recommendation 2

The analysis provides evidence of high fatigue at TOD during late finish FDPs. It is recommended to require operators to apply appropriate fatigue risk management to mitigate the fatiguing effect of late finish FDPs, regardless of FDP duration.

Recommendation 3

The analysis provides evidence of high fatigue at TOD during night FDPs of both long duration (> 10h) and shorter duration (≤ 10h). It is recommended to require operators to apply appropriate fatigue risk management to mitigate the fatiguing effect of *all* night FDPs, regardless of FDP duration.

Recommendation 4

Within night FDPs, duty periods that end at 6.00 or later combined with a start at 1.59 or earlier show the greatest probability of high fatigue at TOD. It is recommended that the regulation define this category of FDP and require operators to pay specific attention to these FDPs when applying fatigue risk management for *all* night FDPs, as proposed in recommendation 3.

Recommendation 5

The analysis found shorter prior sleep to be a predictor of high fatigue at TOD for all night FDPs. The current guidance material for night duties (GM1 CS FTL.1.205) stipulates that it is 'critical for the crew member to obtain sufficient sleep' for night duties longer than 10 hours. It is recommended that the GM be amended to state that it is critical for the crew member to obtain sufficient sleep before *all* night duties, regardless of FDP duration.

Recommendation 6

The analysis provides evidence of high fatigue at TOD during night FDPs. This phenomenon seems to be fairly independent of FDP characteristics (e.g. start and end times, duration), as long as the FDP in question meets the criteria for a night FDP. Prior sleep is the main predictor of eventual fatigue, therefore we recommend that for night FDPs, operators should be required to promote optimum use of sleep opportunities (i.e. before reporting and during the FDP).

How to implement recommendations 5 and 6?

Our research found that shorter sleep in the 24 hours prior to TOD is a predictor of high fatigue at TOD for all night FDPs. Ensuring that crew members obtain sufficient sleep is a shared responsibility of the operator and the crew. Current regulations (ORO.FTL.110 and 115) already describe the need for operators to provide resting opportunities and for crew to make optimum use of such opportunities. As this is essential for effective fatigue risk management, EASA and national civil aviation authorities may need to promote more decisively the use of resting opportunities. To implement this recommendation, we suggest addressing the practicalities conducive to attaining sleep prior to reporting or during the FDP. Providing rest facilities for crew members at or near the airport would improve the probability of obtaining sleep as close as possible to the start of the night duty (as referred to in GM1 CS FTL.1.205 (a)(2) FDP on appropriate fatigue risk management for night duties). That might imply providing suitable accommodation (as defined in ORO.FTL.105 Definitions) at the reporting point for napping in the afternoon prior to a night duty, e.g. for commuting crew members and during the FDP when crew members are on the ground, such as during a long turnaround.

A way of improving opportunities for *in-flight* sleep is the use of an augmented crew (only applicable for longer flights). On augmented flights, the resting crew member uses a seat in the cabin (that meets regulatory requirements) or a separate rest facility. Van den Berg et al. (2016) and Gander et al. (2013) showed that additional in-flight sleep effectively mitigated fatigue on longer flights.

As fatigue can sometimes occur unexpectedly, the use of controlled rest could be considered. 'Controlled rest' means an 'off task' period of time that may include actual sleep. Research has shown that short, controlled rest is an effective in-flight fatigue-mitigation strategy[34], as it can enhance alertness and performance. Our data shows that the napping frequency is higher than expected during night duties. According to GM1 CAT.OP.MPA.201, controlled rest is not pro-active fatigue management (i.e. which is planned before the flight) and may be performed only to manage unexpected fatigue and to reduce the risk of fatigue during higher workload periods later in the flight[35].

We suggest promoting the development and use of controlled rest procedures to enable pilots and cabin crew to take a nap during night FDPs to manage unexpected fatigue and to reduce the risk of fatigue during higher workload periods later in the flight. Pilots can take their controlled rest on the flight deck, whereas cabin crew would have to be provided with suitable bunks or seats by the operator. We also suggest that operators track the use of controlled rest as it is a very useful indication of where additional more effective controls may be necessary.

Follow-on actions

Areas for improvement of the current FTL research and requirements for additional research to be carried out are listed below.

Increase the number of participants and resulting FDPs for consecutive disruptive schedules

We did not achieve the required sample sizes[36] to evaluate more than two consecutive disruptive FDPs of any type. In addition, for two consecutive late duties and mixes, it was not clear whether fatigue increased from one duty to the next. This limited our analyses.

The sample size for consecutive disruptive (i.e. more than two and mixes of disruptive schedules) FDPs could be improved by performing 'more controlled' data collection, compared with the crowdsourcing-like data collection method we used to recruit volunteers. This could be achieved, for example, by requesting the schedules from the airlines beforehand and specifically recruiting crew members operating the consecutive disruptive patterns of interest. Another manner to improve the sample size for consecutive FDPs would be to extend the data collection period beyond two weeks.

Study the remaining four FDPs of interest

The overall FTL review was broken down into smaller research phases, each focusing on specific FDPs. The first and current research phase studied the two top ranked FDPs[37]; i.e. night FDPs of more than 10 hours and disruptive schedules.

The FDPs ranked from third to sixth based on their expected level of fatigue, listed below, still remain to be studied.

- Ranked 3: duties of more than 11 hours for crew members in an unknown state of acclimatisation;

[34] Rosekind et al. 1994
[35] ICAO Fatigue Management Guide for Operators, 2015
[36] As established in D2.2 (Definition of the Data Collection Process) in chapter 2, page 12.
[37] As ranked in D1 Addendum in chapter 3 (page 9) using bio-mathematical modelling and an online survey.

- Ranked 4: duties including a high level of sectors (more than six);
- Ranked 5: duties of more than 13 hours at the most favourable time of the day; and
- Ranked 6: on-call duties such as standby or reserve, followed by flight duties.

The specific FTLs that apply to these six duty period types are generally intertwined with one another. It would therefore be preferable that the recommendations resulting from studies on all six FDPs of interest are available and to use these together to update the FTL rules, if necessary.

Study the effectiveness of the fatigue management training programme

Training and education are fundamental requirements for fatigue risk management. However, few studies have evaluated the short- and long-term effects of such programmes. A recent study by Pylkkönen et al. (in press) failed to find evidence that non-recurrent alertness-management training was an effective remedy for driver sleepiness in occupational settings. These findings conflict with previous studies reporting positive effects of training on alertness management[38], sleep quality[39], and on-duty sleepiness[40]. Contrary to Pylkkönen et al. (in press), these studies evaluated intervention effects using retrospective self-report questionnaires on sleep and on-duty alertness, before and after an intervention, without actual field measurements.

The AMC1 ORO.FTL.250 Fatigue Management Training provides details on content to be covered in initial and recurrent fatigue management training for crew members, for personnel responsible for the preparation and maintenance of crew rosters, and for the management personnel concerned. It would be worthwhile to study further the effectiveness of the fatigue management training content and frequency.

Look further into fatigue mitigation measures focusing specifically on more sleep (opportunity)

As prior sleep (opportunity) is such a strong determinant of fatigue, we suggest studying effective ways of achieving more sleep. A few potentially interesting research directions are discussed below, all of which warrant further investigation to better assess their operational feasibility and effectiveness.

The first area concerns the effectiveness of in-flight sleep, in the cabin or bunk. Previous studies suggest that in-flight sleep obtained in crew rest facilities (horizontal sleeping position) is lighter and more fragmented than sleep in a layover hotel[41]. Other studies have found that pilots have difficulty converting time in on-board rest facilities into a reasonable amount of sleep[42]. In addition, sleep inertia is a factor to be reckoned with when designing in-flight sleep strategies, and thus also deserves attention.

A second area concerns the operational feasibility and effectiveness of napping on the flight deck or in the cabin. Our data shows that the napping frequency is higher than expected as napping on the flight deck is only allowed to manage unexpected fatigue and to reduce the risk of fatigue during higher workload periods later in the flight. Therefore, we suggest examining the planned use of in-flight napping as a strategic measure.

[38] Gander et al. 2005; Rosekind et al. 2006
[39] Nishinoue et al. 2012
[40] Kakinuma et al. 2010; Van Drongelen et al. 2014
[41] Ho et al. 2005; Signal et al. 2005, 2013
[42] Signal et al. 2005; Roach et al. 2011

General scheduling principles based on fatigue science are provided in GM1 CS FTL.1.205 (a) (2) FDP and the Manual for the Oversight of Fatigue Management Approaches (ICAO Doc 9966), which ORO.FTL refers to for further guidance on fatigue risk management. These should be taken into account when designing duty rosters. However, we expect that adding real examples from different types of operations will help operators and schedulers to ensure adequate sleep for crews prior to a night FDP.

A particular aspect of rostering that our survey revealed and that could be interesting to examine further is roster changes. Survey respondents[43] ranked 'roster changes' as the second (out of 11) most frequent factor affecting their level of fatigue in the upcoming duty. This is likely because changes, which may well be implemented at the last minute, make it more difficult for crew members to achieve sufficient sleep before their duty commences.

In the context of fatigue management training, a promising option to achieve more sleep prior to night flights is personalised and context-specific interventions. For instance, tailored advice regarding exposure to daylight, sleep, physical activity and nutrition could be provided, perhaps using mobile health technology[44].

Look for an alternative to PVT as an objective performance measure

The use of PVT as an objective performance measure was unsuccessful in the current field study set-up. Pilots were asked to perform the PVT 15 minutes prior to TOD in their final sector. Note that cabin crew were *not* asked to perform the PVT, as we expected that it would be difficult for them to do so undisturbed. Pilots stated numerous times that the PVT was too burdensome and disturbing to perform on the flight deck. In addition, the pilots did not always perform the PVT in line with the protocol, which stated that the test should be done 15 minutes prior to the TOD of the final sector of that day. This resulted in inadequate data points for inclusion of PVT data in the final data set.

Many studies have successfully collected PVT data in flight operations under more *controlled* conditions as compared to the approach used for the current study[45]. The reason for our lack of success could be that no researcher was present to make sure the test was performed in line with the study protocol and without disturbances, or that we did not provide the crew with *individual* training on the PVT. Nonetheless, as fatigue risk management processes may also require collecting data, operators would benefit from an alternative to PVT to objectively measure performance (i.e. alertness). Alternatives should be less burdensome and intrusive and less sensitive to disturbances. Additionally, it would be valuable if future objective performance measurement methods could record fatigue-related data in a continuous manner. This would allow the assessment of the length of time that crew actually are fatigued and thus improve the identification and assessment of fatigue hazard.

Study of diet and physical exercise as countermeasures to fatigue

ORO.FRL.240 on nutrition describes regulatory requirements on opportunities to eat and drink to avoid dips in crew members' performance. Indeed, prior research has found reduced meal frequency among shift workers, but snacking is more common on the night shift[46]. Night work has also been found to reduce opportunities for physical activity and participation in sports[47]. Optimal timing of physical activity and intake of

---

[43] As presented in D2.1 (Identification of Potential Fatigue Hotspots) in chapter 3, page 14.
[44] Van Drongelen et al. 2014; Lentferink et al. 2017
[45] As listed in D1 (Definition of the Baseline) in chapter 2 (page 6).
[46] Atkinson et al. 2008
[47] Waterhouse et al. 2007

specific nutrients can enhance sleep duration and quality, thus stimulating alertness or relaxation[48]. Access to suitable catering facilities that provide nutritional food and beverages, the consistent availability of advice on diet, and performance of physical exercise are areas perhaps insufficiently promoted as effective countermeasures to fatigue. Survey respondents[49] ranked 'poor nutrition' as seventh (out of 10) most frequent conditions that worsened their level of fatigue. It might be worthwhile to review the state of the art on diet and physical exercise and extend the work that has been done to come up with best practices for this.

[48] Waterhouse et al. 2007; Atkinson et al. 2008; Fischer et al. 2002
[49] As presented in D2.1 (Identification of Potential Fatigue Hotspots) in chapter 3, page 15.

Chapter 8: Critique of the whole data collection activity

This chapter looks critically at the data collection process and outcomes, including factors that that may have adversely impacted on the size of the sample and the quality of the data. Upon completion of the data collection phase, we held debriefings with airline coordinators, and they provided inputs for this chapter. In addition, we spoke with other airline personnel, the project manager, and consortium members. In particular, we asked them what factors, in their view, may have impacted on the scale and quality of the data collected. We also considered feedback received from participants via telephone and email.

Critique assessment

The current data collection was unique in that 24 different airlines agreed to participate and a large number of crew members within Europe were invited to join the field study and gather data. This method yielded a sample over which the project only had indirect control. The control that we did have was via the airline coordinators, who acted as liaisons to their airline and crew members.

Due to our crowdsourcing-based participant recruitment method, we also lacked the ability to control adherence to the measurement protocol. The crew members were offered training material to familiarise and train themselves in the use of the protocol and the app for data collection. In addition, we explained the details of the data collection procedure to the airline coordinators, so that they could answer questions from crew members. However, we could not be sure if and for how long this material was studied and used. We could only ensure that the volunteers had easy access to the material and ample opportunity to ask questions to either the airline coordinator or the principal investigator. A dedicated website was created with information about the project and promotion and training material was sent to the airline coordinators for the crew rooms. Informational emails were sent to the volunteers who registered to participate via the website.

The airline coordinators communicated to crew members in their native language in most cases, which worked well, according to the airline coordinators. However, once the volunteers clicked on the NLR web link provided on the invitation mailing for follow-up information and registration, the content was only provided in English. This turned out to be an issue for some participants. In particular, the training module included a short technical explanation of how to work with the app, which some crew said it was difficult to understand.

The app that was used only runs on Apple devices, which narrowed down the population of interest. We received some emails from volunteers stating they could not participate due to this limitation.

The subsequent research on the remaining four types of duty periods should undertake specific actions to increase the participation of women and cabin crew. The current field study population consisted of 75 % male and 68 % pilots.

References

Åkerstedt, T. (1988). Sleepiness as a consequence of shift work. *Sleep, 11(1)*, 17-34.

Åkerstedt, T., Anund, A., Axelsson, J., & Kecklund, G. (2014). Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *Journal of Sleep Research, 23(3)*, 240-252.

Atkinson, G., Fullick, S., Grindey, C., Maclaren, D., & Waterhouse, J. (2008). Exercise, energy balance and the shift worker. *Sports Medicine, 38(8)*, 671-685.

Dawson, D., & McCulloch, K. (2005). Managing fatigue: it's about sleep. *Sleep Medicine Reviews, 9(5)*, 365-380.

Fischer, K., Colombani, P. C., Langhans, W., & Wenk, C. (2002). Carbohydrate to protein ratio in food and cognitive performance in the morning. *Physiology & Behavior, 75(3)*, 411-423.

Folkard, S., & Åkerstedt, T. (1997). The three-process model of alertness and its extension to performance, sleep latency, and sleep length. *Chronobiology International, 14(2)*, 115-123.

Gander, P. H. (2015). Evolving regulatory approaches for managing fatigue risk in transport operations. *Reviews of Human Factors and Ergonomics, 10(1)*, 253-271.

Gander, P. H., Marshall, N. S., Bolger, W., & Girling, I. (2005). An evaluation of driver training as a fatigue countermeasure. *Transportation Research Part F: Traffic Psychology and Behaviour, 8(1)*, 47-58.

Gander, P. H., Signal, T. L., 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*, 697-706.

Ho, P., Landsberger, S., Signal, L., Singh, J., & Stone, B. (2005). The Singapore experience: task force studies scientific data to assess flight. *Flight Safety Digest, 26*, 20-40.

Kakinuma, M., Takahashi, M., Kato, N., Aratake, Y., Watanabe, M., Ishikawa, Y., Kojima, R., Shibaoka, M., & Tanaka, K. (2010). Effect of brief sleep hygiene education for workers of an information technology company. *Industrial Health, 48(6)*, 758-765.

Lentferink, A. J., Oldenhuis, H., K., De Groot, M., Polstra, L., Velthuijsen, H., & Van Gemert-Pijnen, J. E. (2017). Key components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: a scoping review. *Journal of Medical Internet Research, 19(8)*, e277.

Nishinoue, N., Takano, T., Kaku, A., Eto, R., Kato, N., Ono, Y., & Tanaka, K. (2012). Effects of sleep hygiene education and behavioral therapy on sleep quality of white-collar workers: a randomized controlled trial. *Industrial Health, 50(2)*, 123-131.

Monk, T, H. (1990). Shiftworker performance. *Occupational Medicine. 5(2)*, 183-198.

Pylkkönen, M., Tolvanen, A., Hublin, C., Kaartinen, J., Karhula, K., Puttonen, S., Sihvola, M., Sallinen, M. (in press). Effects of alertness management training on sleepiness among long-haul truck drivers: a randomized controlled trial. *Accident Analysis and Prevention*.

Rosekind, M. R., Graeber, R. C., Dinges, D. F., Connell, L. J., Rountree, M. S., Spinweber, C. L., & Gillen. K. A. (2014). Crew Factors in Flight Operations IX: Effects of Planned Cockpit Rest on Crew Performance and Alertness in Long-Haul Operations (NASA Technical Memorandum No 108839). Moffett Field: National Aeronautics and Space Administration.

Rosekind, M. R., Gregory, K. B., & Mallis, M. M. (2006). Alertness management in aviation operations: Enhancing performance and sleep. *Aviation Space and Environmental Medicine, 77(12)*, 1256-1265.

Sallinen, M., Åkerstedt, T., Härmä, M., Henelius, A., Ketola, K., Leinikka, M., Kecklund, G., Sihvola, M., Tuori, A., Virkkala, J., & Puttonen, S. (2018). Recurrent On-Duty Sleepiness and Alertness Management Strategies in Long-Haul Airline Pilots. *Aerospace Medicine and Human Performance, 89(7)*, 601-608.

Sallinen, M., & Hublin, C. (2015). Fatigue-Inducing Factors in Transportation Operators. In: Popkin S. M. (ed.), *Reviews of Human Factors and Ergonomics: Worker Fatigue and Transportation Safety*, vol. 10, pp. 138-167. Thousand Oaks: SAGE Publications.

Signal, T. L., Gale, J., & Gander, P. H. (2005). Sleep measurement in flight crew: comparing actigraphic and subjective estimates of sleep with polysomnography. *Aviation, Space, and Environmental Medicine, 76*, 1058-1063.

Signal, T. L., Gander, P. H., Van den Berg, M. J., & Graeber, R. C. (2013). In-flight sleep of pilots during a 7-hour rest break: implications for research and flight safety. *Sleep, 36*, 109-115.

Van den Berg, M. J., Wu, L. J., & Gander, P. H. (2016). Subjective measurement of in-flight sleep, circadian variation, and their relationship with fatigue. *Aerospace Medicine and Human Performance, 87(10)*, 869-875.

Van Drongelen, A., Boot, C. R., Hlobil, H., Twisk, J. W., Smid, T., & Van der Beek, A. J. (2014). Evaluation of an mHealth intervention aiming to improve health-related behavior and sleep and reduce fatigue among airline pilots. *Scandinavian Journal of Work, Environment & Health, 40(6)*, 557-568.

Waterhouse, J., Reilly, T., Atkinson, G., & Edwards, B. (2007). Jet lag: trends and coping strategies. *Lancet, 369(9567)*, 1117-1129.

Wesensten, N. J., Balkin, T. J., & Belenky, G. (2015). Countermeasures for mitigating fatigue in motor vehicle operators. In: Stephen M. Popkin (ed.), *Reviews of Human Factors and Ergonomics: Worker Fatigue and Transportation Safety*, vol. 10, pp. 115-137. Thousand Oaks: SAGE Publications.

Among all 24 non-European studies included in the state of the art review (Chapter 2: Baseline for the research), the 15 studies listed below were of extra interest because the design and methods resembled those used in the current FTL research.

1. Berg, M. J. van den, Signal T. L., Mulrine, H. M., Smith, A. A. T., Gander, P. H., & Serfontein, W. (2015). Monitoring and managing cabin crew sleep and fatigue during an ultra-long range trip. *Aerospace Medicine and Human Performance*, *86(8)*, 705-713.
2. 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.
3. Gander, P. H., Signal, T. L., Berg, M. J., Mulrine, H. M., Jay, S. M., & Mangie, J. (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.
4. Greeley, H. P., Roma, P. G., Mallis, M. M., Hursh, S. R., Mead, A. M., & Nesthus, T. E. (2013). Field study evaluation of cepstrum coefficient speech analysis for fatigue in aviation cabin crew. FAA Report No DOT/FAA/AM-13/19. Washington, DC: Federal Aviation Administration.
5. Holmes, A., Al-Bayat, S., Hilditch, C., & Bourgeois-Bougrine, S. (2012). Sleep and sleepiness during an ultra long-range flight operation between the Middle East and United States. *Accident Analysis & Prevention*, *45*, 27-31.
6. Mello, M. T. de, Esteves, A. M., Pires, M. L. N., Santos, D. C., Bittencourt, L. R. A., Silva, R. S., & Tufik, S. (2008). Relationship between Brazilian airline pilot errors and time of day. *Brazilian Journal of Medical and Biological Research*, *41(12)*, 1129-1131.
7. 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*, 1013-1017.
8. Powell, D., Spencer, M. B., Holland, D., & Petrie, K. J. (2008). Fatigue in two-pilot operations: implications for flight and duty time limitations. *Aviation, Space, and Environmental Medicine*, *79(11)*, 1047-1050.
9. Powell, D., Spencer, M. B., Holland, D., Broadkent, E., & Petrie, K. J. (2007). Pilot fatigue in short-haul operations: effects of number of sectors, duty length, and time of day. *Aviation, Space, and Environmental Medicine*, *78*, 698-701.
10. 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.
11. Roach, G. D., Sargent, C., Darwent, D., & Dawson, D. (2012). Duty periods with early start times restrict the amount of sleep obtained by short-haul airline pilots. *Accident Analysis & Prevention*, *45*, 22-26.
12. 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*, 1199-1208.
13. Thomas, M. J. W., & Ferguson, S. A. (2010). Prior sleep, prior wake, and crew performance during normal flight operations. *Aviation, Space, and Environmental Medicine*, *81(7)*, 665-670.
14. Thomas, M. J. W., Petrilli, R., M., & Roach, G. D. (2007). The impacts of Australian transcontinental ‘back of clock’ operations on sleep and performance in commercial aviation flight crew. ATSB Report No B2005/0121. Canberra: Australian Transport Safety Bureau.
15. 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. Doctoral dissertation, Washington State University, Pullman.

Definitions

Accommodation means, for the purpose of standby and split duty, a quiet and comfortable place not open to the public with the ability to control light and temperature, equipped with adequate furniture that provides a crew member with the possibility to sleep, with enough capacity to accommodate all crew members present at the same time and with access to food and drink (ORO.FTL.105 (3)).

Augmented flight crew means a flight crew which comprises more than the minimum number required to operate the aircraft, allowing each flight crew member to leave the assigned post, for the purpose of in-flight rest, and to be replaced by another appropriately qualified flight crew member (ORO.FTL.105 (5)).

Disruptive schedule means a crew member's roster which disrupts the sleep opportunity during the optimal sleep time window by comprising an FDP or a combination of FDPs which encroach, start or finish during any portion of the day or of the night where a crew member is acclimatised. A schedule may be disruptive due to early starts, late finishes or night duties (ORO.FTL.105 (8)).

Duty means any task that a crew member performs for the operator, including flight duty, administrative work, giving or receiving training and checking, positioning, and some elements of standby (ORO.FTL.105 (10)).

Duty period means a period which starts when a crew member is required by an operator to report for or to commence a duty and ends when that person is free of all duties, including post-flight duty (ORO.FTL.105 (11)).

Early type of disruptive schedule means (ORO.FTL.105 (8a)):
(i) for 'early start' a duty period starting in the period between 5:00 and 5.59 in the time zone to which a crew member is acclimatised; and
(ii) for 'late finish' a duty period finishing in the period between 23:00 and 1.59 in the time zone to which a crew member is acclimatised.

Late type of disruptive schedule means (ORO.FTL.105 (8b)):
(i) for 'early start' a duty period starting in the period between 5:00 and 6.59 in the time zone to which a crew member is acclimatised; and
(ii) for 'late finish' a duty period finishing in the period between 24.00 and 1.59 in the time zone to which a crew member is acclimatised.

Flight duty period (FDP) means a period that commences when a crew member is required to report for duty, which includes a sector or a series of sectors, and finishes when the aircraft finally comes to rest and the engines are shut down, at the end of the last sector on which the crew member acts as an operating crew member (ORO.FTL.105 (12)).

Night duty means a duty period encroaching any portion of the period between 2.00 and 4:59 in the time zone to which the crew is acclimatised (ORO.FTL.105 (9)).

Rest facility means a bunk or seat with leg and foot support suitable for crew members' sleeping on board an aircraft (ORO.FTL.105 (19)).

Sector means the segment of an FDP between an aircraft first moving for the purpose of taking off until it comes to rest after landing on the designated parking position (ORO.FTL.105 (24)).

Suitable accommodation means, for the purpose of standby, split duty, and rest, a separate room for each crew member located in a quiet environment and equipped with a bed, which is sufficiently ventilated, has a device for regulating temperature and light intensity, and access to food and drink (ORO.FTL.105 (4)).

Window of circadian low (WOCL) means the period between 2.00 and 5.59 hours in the time zone to which a crew member is acclimatised (ORO.FTL.105 (28)).

Deep night FDP: FDPs ending at 6.00 or later and starting at 1.59 or earlier.

Deep early start FDP: FDPs starting between 2.00 and 4.59.

Daytime FDP: all FDPs with start time ≥ 7.00 and end time < 23.00.

Baseline FDP: all collected FDPs.

Abbreviations

Abbreviation	Description
AMC	Acceptable Means of Compliance
BAM	Boeing Alertness Model
CAS	Common Alertness Scale
CAT	Commercial Air Transport
CI	Confidence Interval
CS	Certification Specification
D	Deliverable
EASA	European Aviation Safety Agency
EC	European Commission
EU	European Union
FAST	Fatigue Avoidance Scheduling Tool
FDP	Flight Duty Period
FTL	Flight Time Limitation
GM	Guidance Material
ICAO	International Civil Aviation Organization
KSS	Karolinska Sleepiness Scale
MPA	Motor-Powered Aircraft
OP	Operating Procedures
OR	Odds Ratio
ORO	Organisation Requirements (in the air Operations Regulation)
PVT	Psychomotor Vigilance Task
SAFTE	Sleep, Activity, Fatigue, and Task Effectiveness
SP	Samn-Perelli
TOD	Top Of Descent
WOCL	Window Of Circadian Low

European
Commission