European Union Aviation Safety Agency

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

RESEARCH PROJECT EASA.2019.C31

DELIVERABLE 2.1: DATA COLLECTION SCOPE AND PROCESS

Effectiveness of Flight Time Limitations (FTL 2.0)

Disclaimer

DELIVERABLE NUMBER AND TITLE: **FTL D2.1. Data collection scope and process (final).**

CONTRACT NUMBER: EASA.2019.C31

CONTRACTOR / AUTHOR: NLR / Alwin van Drongelen

IPR OWNER: European Union Aviation Safety Agency

DISTRIBUTION: Restricted

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

DATE: **02 August 2023**

SUMMARY

Problem area

The research study FTL 2.0 aims to perform a 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. More specifically, the main purpose is to add to, and subsequently complete, the work performed during the first phase of the “Effectiveness of Flight Time Limitation” evaluation (MOVE/C2/2016-360).

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

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

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

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

d) On-call duties such as standby or reserve followed by flight duties.

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

Description of work

The current deliverable (D2.1) involves the definition of scope and process for the data collection. The main objective was to define the scope and scale for the data collection and processing and the main protocols and requirements that rule such processes. This document thus delivered a scientifically based, as well as practically executable, fatigue measurement methodology that will be applied in the FTL 2.0 study.

Results and Application

The starting point for the methodology described was the proposal for this project, that in turn was based on the methodology that was applied in the previous FTL project. However, the current study will use an extended and in some respect different set of measurement techniques where needed, and data collection procedures that allow for more high quality data. For all measurement instruments considered, their relevance to the duties of interest, their scientific value as well as practical aspects were considered. The outcomes of the selection procedure are described in Chapter 2.

The work performed also led to a revision of the proposed data collection procedure and the study designs used to evaluate the effectiveness of the FDPs of interest. The revised data collection procedure can be found in Chapter 0. The different revised study designs are subsequently described in Chapter 4.

Finally in Chapter 5, the data collection governance scheme which is in accordance with the data collection approach presented in the previous chapters is described, ensuring the quality of the data to be collected and the protection of the data of the people involved.

CONTENTS

ABBREVIATIONS

ACRONYM	DESCRIPTION
COTS	Commercial off The Shelf
CR	Controlled Rest
D	Deliverable
DLMO	Dim Light Melatonin Onset
DLR	German Aerospace Centre
EASA	European Union Aviation Safety Agency
ECG	Electrocardiogram
EEG	Electroencephalography
EOG	Electrooculography
FDP	Flight Duty Period
fNIRS	Functional near-infrared spectroscopy
FRM	Fatigue Risk Management
FTL	Flight Time Limitations and rest requirements
FTT	Finger Tapping Test
HRV	Heart Rate Variability
ICAO	International Civil Aviation Organization
ID	Identification
ISA	Instantaneous Self-Assessment
KSS	Karolinska Sleepiness Scale
MCTQ	Munich Chronotype Questionnaire
NASA TLX	NASA Task Load Index
NLR	Royal Netherlands Aerospace Centre
PM	Project Manager
PMP	Project Management Plan
PoC	Point of Contact
PSG	polysomnography
PVT	Psychomotor Vigilance Task
rMEQ	Reduced Morningness-Eveningness questionnaire
RSME	Rating Scale Mental Effort
SCN	suprachiasmatic nucleus
SP	Samn Perelli
SSL	Secure Sockets Layer
ToC	Top of Climb
ToD	Top of Descent
TST	Total Sleep Time
VAS	Visual Analogue Scale
WOCL	Window of Circadian Low

1. Introduction

1.1 Project description

The main objective of this Effectiveness of Flight Time Limitation (EASA.2019.C31) study, also know as FTL 2.0, is to perform a 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. More specifically, the main purpose is to add to the work performed during the first phase of the "Effectiveness of Flight Time Limitation" evaluation (MOVE/C2/2016-360) and complete the assessments of the aspects of the FTL rules identified by EASA.

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

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

Two main series of deliverables will be provided in this study:

- D1: deliverables on the work performed in Task 1.1 and Task 1.2: this deliverable is divided in two sub-deliverables: the definition of baseline (1.1) and the definition of the target crew population (1.2);
- D2: deliverables on the work performed in Task 2.1 to Task 2.7: this deliverable is divided in four sub-deliverables: the definition of scope and process for the data collection (Task 2.1), the data repository (Task 2.2), the data analyses and benchmark against other reference sources (Task 2.3 and 2.4), and the synopsis of the results of the previous and current contract including a list of generally applicable performance metrics, the analysis of suitability of the examined fatigue management measures, and the conclusions and recommendations (Task 2.5, 2.6 and 2.7).

1.2 Parties involved

To execute the research study a consortium is formed that is led by the Royal Netherlands Aerospace Centre (NLR), furthermore consisting of Stockholm University (SU), the Finnish Institute of Occupational Health (FIOH), German Aerospace Centre (DLR) and Jeppesen.

Royal Netherlands Aerospace Centre (NLR)
Project Manager and Site Manager:
Dr. Alwin van Drongelen

Stockholm University (SU)
Site Manager:
Dr. Torbjörn Akerstedt

Finnish Institute of Occupational Health (FIOH)
Site Manager:
Dr. Mikael Sallinen

German Aerospace Centre (DLR)
Site Manager:
Dr. Daniel Aeschbach

Jeppesen
Site Manager:
Mr. Tomas Klemets

The contracting authority EASA is represented by a Project Manager and a Technical Lead who both serve as contact points during the duration of the contract.

Project Manager:
Emmanuel Isambert

Technical Lead:
Irina Petrova

1.2.1 Mirror Group

A "Mirror group" of representatives from the main interested parties – Member States' competent authorities, airlines and aircrew associations – was set-up by EASA after the start of the project. The Mirror Group has the following roles:

- To oversight the definition of the scope and scale of the work to be performed with the goal of guaranteeing its fitness for purpose within the European aviation sector; d
- To advise and facilitate the required interactions with third parties that potentially need to be involved in the project;
- To assist in arranging the airline/aircrew participation for the data collection part of the project.

The Mirror Group gets updated on the progress of the project every four to six months.

1.2.2 Scientific Committee

A committee of independent scientific experts was set-up by EASA for purposes of providing assistance with the implementation of the research activities. This group of experts was notably assigned to validate specific scientific or technical approaches that are to underpin the work as well as to peer-review the technical and final deliverables of the contract. The members of the Scientific Committee are:

- Dr. Alex Holmes (Clockwork Research)
- Dr. Barbara Stone (FRMSC)
- Cpt. Kristjof Tritschler (smartshiftwork.com / Lufthansa)

1.3 This deliverable

The current deliverable involves the definition of scope and process for the data collection. The main objective is to define the scope and scale for the data collection and processing and the main protocols and requirements that rule such processes. This deliverable will include a description of:

1. The scope and timeline of the data collection and processing, including notably the identification of the spectrum of research parameters to be collected, together with the rationale justifying their selection; and
2. The specification of the data collection, processing, and protection protocols with the relevant support documentation – e.g. data collection and processing best-practice guidelines and standards, and quality management.

Thus, this document delivers a scientifically based, as well as practically executable, fatigue measurement methodology that will be applied in the current FTL study (FTL 2.0). It supports and justifies the choices that were made when selecting the measurement tools and methods. The proposed methodology was carefully considered to meet current scientific standards while remaining a practical methodology for use in an operational setting and within the available constraints (time and budget).

1.4 Approach

The methodology for this project, takes into account the methodology applied in the previous FTL project, thus allowing for a compatibility and comparison between the results of this study (FTL2.0) and those of the earlier study (section 2.2). However, where needed, the current study will use an extended and in some respect different set of measurement techniques, and data collection procedures that allow for more high quality data.

The data collection in FTL2.0 uses the outcomes of Task 1.1 (review of the state of the art) as a basis. The literature search of Task 1.1 resulted in a substantial amount of articles concerning the measurement of fatigue, sleep and workload in the aviation domain. For all measurement instruments considered in these papers, the relevance to the duties of interest, scientific value as well as practical aspects were considered. A description of the measurement techniques considered can be found in Annex A.

Chapter 2 describes the outcomes of the selection procedure of the most eligible techniques and explaines in what way the measurement approach deviates from the project proposal. Chapter 2 also includes a table comparing the methodology to be used in this project and the one used in the previous contract.

The results of the literature review and interviews performed in Task 1.1, together with extensive internal and external (with EASA, the Scientific Committee, and the Mirror Group) discussions of the consortium, also led to a revision of the proposed data collection procedure and the study designs used to evaluate the effectiveness of the FDPs of interest in this project. The revised data collection procedure can be found in Chapter 0. The corresponding revised study designs are subsequently described in Chapter 4. Finally, in Chapter 5, the data collection governance scheme that is set up to ensure high quality data and the protection of the data of the people involved, is described.

2. State of the art of measurement approach

In this chapter, the research parameters are defined and justified in accordance with the scope of the data collection. For this, a literature-based review of methods to measure fatigue, sleep, workload, and related factors in the field was performed (outcomes provided in Annex A). Based on this review, the most appropriate combination of measures was selected, based on validity, operational feasibility, and lessons learned from the previous contract. This procedure led to a revised set of selected measurement techniques for the data collection phase of the FTL 2.0 project in comparison with the project proposal. At the end of this chapter, a table is provided in which the measurement approach differences between the previous contract (FTL) and the current project (FTL2.0) can be found.

2.1 Selected measurement techniques

Since data collection will be done in the field, the proposed methods for this project must be practical in use, and be applicable in an unsupervised setting during flight operations. The subjects will be aircrew, both pilots on the flight deck carrying out their normal duties, and cabin crew who also have to perform a number of safety related and other tasks such as manoeuvering trolleys in small-gallies and aisles. As such, the instruments that will be applied must be non-intrusive, robust, permitted by regulation, and able to record and store data without the assistance of a researcher. Furthermore, the study needs to be executed on a large scale and limited timeframe for which self-administration (usage by crew in-flight and on the flight deck) is important. As such, it would not be feasible to include measurement devices that require researchers to travel along to equip the participants and calibrate the devices. Valid electroencephalography (EEG) measurements for instance would provide valuable data, but would also be unacceptably intrusive and would require a researcher to apply and calibrate the measurement devices.

We therefore aimed to select the best set of measures that is feasible in practice but yet provides a robust assessment of aircrew fatigue, workload, and sleep. The selection made is based on criteria determined by the consortium members of this project.The measurement or combination of measurements should:

- be fast and easy to undertake in an operational environment;
- allow for self-administration;
- be robust against falsification of parameters;
- only produce personal data necessary for the purpose of the research;
- provide results that unambiguously indicate the level of fatigue, workload, and sleep;
- be socially acceptable and non-invasive; and
- provide results that are not subject to learning effects.

2.1.1 Measurement devices

The selected measurement devices are:

1. Research-grade wrist-worn actigraphy monitors by Philips Respironics (AW2 and Spectrum Plus) (Philips Respironics, Inc.; Murrysville, PA, USA). The project team has 37 AW2 actigraphs, and 25 Spectrum Plus actigraphs that can be used throughout the whole data collection phase. For the specific FDP on duties of more than 11 hours for crew members in an unknown state of acclimatisation, 20 additional Spectrum Plus actigraphs will be available, since they measure not only light intensity (in lux) but also spectral light composition in three channels (red, blue, green). The latter is required for the mathematical model applied in this FDP.
2. Designated mobile phones which are equipped with the Fit4Duty data collection app developed by DLR. The app includes the 3-min PVT (Psychomotor Vigilance Task) version (for description and validation, see Benderoth et al., 2021), and the phones have been selected to be able to reduce latencies and standard deviations to acceptable thresholds (as defined in Basner et al., 2020), tested against light sensor measurements at DLR. The phones are of the following types: ASUS ROG Phone 2, Nokia 5.4, and Xiaomi Mi 10. The app measures at least the same variables as the personal iphone

based CrewAlert app during the previous contract. Due to the fact that participants receive a specific phone in this study, more crew members can be included in comparison with the previous FTL study (in which only iOS users could be included).

In total, 60 phones will be available for the data collection phase. The Fit4Duty app includes a sleep-wake log, a flight schedule and work-related aspects log, and provides possibilities for subjective fatigue/sleepiness ratings. Furthermore, the PVT can be administered through this app, and the usage of a dedicated mobile phone assures that the recorded reaction times are not biased by the internal memory capacity of the phone and therefore suitable for within-subject comparison (Arthurs et al., 2021). In addition to the ability to accurately record the input of the participants, the app has the following specifications:

- It is user-friendly;
- New types of measures can be added easily;
- Measures can be configured per sub-group of participants;
- The incoming data can be easily monitored through the backend of the app;
- It has an encrypted connection to a secured database where the data is stored; and
- It has the possibility to directly prompt participants when expected data is lacking.

Some screenshots/photos of the app can be found in Figure 2-1 and Figure 2-2.

Before starting with the full-size data collection campaign (Chapter 4) we will perform a trial study using a small group of aircrew members for a period of around three days to test how they handle the equipment. The trial study will be executed within one airline. Lessons learned from this pilot will be taken into account and specific emphasis will be given to the user friendliness of the data collection app and the ability to comply with the protocol as requested (e.g. the burden of filling out the questions, and the frequency and timing of fatigue ratings).

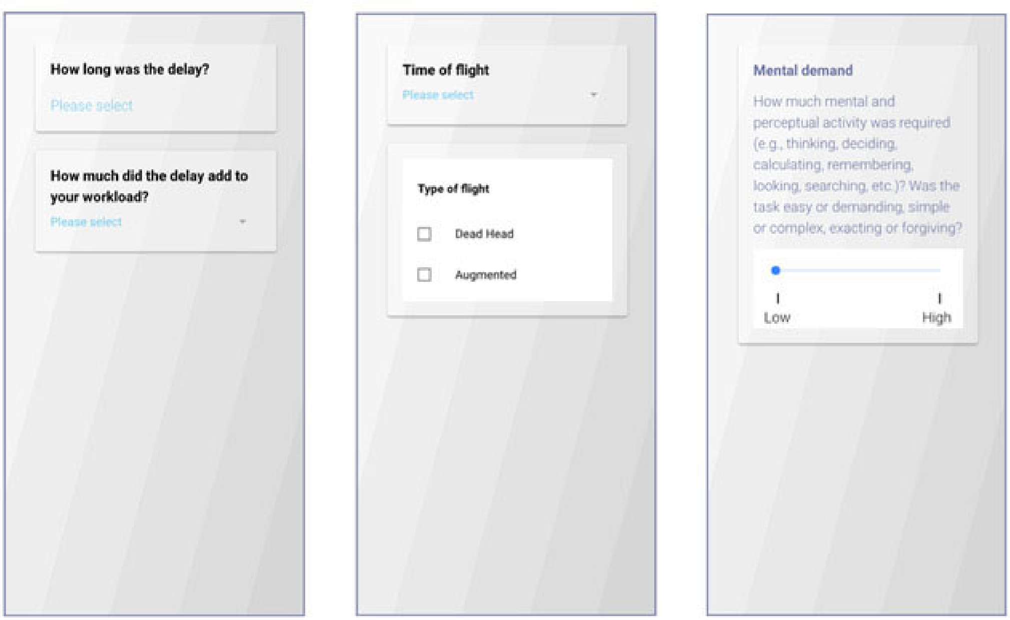

Figure 2-1. Questionnaire examples of the Fit4Duty app (taken from a previous DLR study).

Figure 2-2. The Psychomotor Vigilance Test (PVT) on an ASUS ROG Phone 2.

2.1.2 Data collection measures

Fatigue

The Karolinska Sleepiness Scale (KSS) will be applied to measure sleepiness (Åkerstedt and Gillberg, 1990). It is easy to use in an operational environment and correlates well with decisions by crew to report fatigue. It is also relatively easy to collect and analyse when integrated in the to-be-used data collection app.

The Samn-Perelli (SP) will be measured as well. Although the KSS and SP are often referred to as interchangeable and are highly correlated, they still measure different concepts, i.e., sleepiness (KSS) versus fatigue (SP), and thus provide added value. Since both measures are widely used in aviation research, taking the SP into account will also improve the ability to compare the outcomes of this project with preceding research.

Performance

Performance as an indicator of fatigue will be measured by applying the shortest validated version of the PVT. This 3-minute long version has been developed and validated by DLR (Elmenhorst et al., 2012; Benderoth et al., 2021), and in combination with properly instructing individual participants, it is considered to be a very suitable alternative in comparison with the 5-minute PVT applied in the previous contract. The PVT will be integrated in the Fit4Duty app, which makes it relatively easy to collect and analyze.

Sleep duration and efficiency

Sleep duration and sleep efficiency will be measured objectively by using the research-grade wrist-worn actigraphs described above. These devices have been widely used in sleep research and validation studies, showing similar or better specificity and sensitivity values compared to consumer-graded devices (Cheung et al.,2020). In addition, these ActiWatches have a relative long battery life and can therefore be used for data collection periods of up to 60 days (Spectrum Plus). They also measure data without interference of algorithm-based apps or cloud services, for which raw data can be gathered and data privacy for the participants can be assured. Furthermore, since wrist-worn actigraphy devices do not disrupt aircrew during work, they are a practical and feasible technique to gather relevant sleep related data in addition to subjective sleep logs.

Subjective sleep logs will be used to track sleep/wake cycles. This technique is able to provide reliable estimates of sleep quantity and quality and has often been used in field studies with aircrew before (e.g. Åkerstedt et al., 2021; Petrilli et al., 2006). Furthermore, because the questions will be included in the data collection app, participants can fill them out at a convenient moment, while for the researchers it is easy to extract and analyse the data. The sleep log in this study will comprise questions on sleep timing, time awake, wake-up time, sleep quality, feelings of restedness, and sleep location.

Workload

Workload will be measured by means of two subjective and one objective measure.

After each duty day, the Rating Scale Mental Effort (RSME) will be applied in the data collection app, to rate the peak level of mental effort. Similar to the previous FTL study, participants will be led to an additional screen with possible 'hassle factors' if they slide the RSME scale to a value of 50 points or higher. Based on expert discussions and findings from previous projects, the following 24 hassle factors can be selected by the participants:

- no break
- bad weather
- demanding airport
- high density airspace
- sluggish ground handling
- tight crew rotation
- technical defect
- abnormal procedures
- duty change on short notice
- hotel (noisy/low quality)
- difficult passengers
- low quality food
- long travel from airport to hotel
- short turn around
- delay/time pressure
- emergency
- critical fuel status
- waiting times between flights
- difficulty getting through security
- insufficient baggage handling
- special/challenging airport
- late slot
- runway change
- other

In addition to the RSME, the NASA Task Load index (TLX) will be applied after each duty. Participants are asked to rate all six scales by means of a slider ranging from 'very low' to 'very high'. From these ratings, three outcomes will be determined: mental demand, physical demand, and the mean (unweighted) overall score called 'Raw TLX' (Hart, 2006).

Finally, physical workload will be objectively measured by means of recording of physical activity using the ActiWatches. This is particularly important to get a good indication of the physical workload experienced by cabin crew (which will be higher in comparison with pilots), while it requires no extra effort from their side since they already wear the actigraphs for the objective sleep/wake data collection.

Work-related inputs

Operational flight related data will be gathered from the participants themselves, when asked to log their schedule details, including relevant work-related data such as:

- Departure and arrival times and locations;
- Number of sectors flown;
- Time zone change; and
- Crew composition.

Circadian timing

It is well established that the effects of time zone changes on circadian processes are significant after the crossing of three or more time zones (Samel et al., 1995). It has also become clear however that imposing a one hour time difference (e.g. through the introduction of day light saving time) can lead to sleep disruption and a change in health and regulatory behaviours (Harrison, 2013). This is relevant for aircrew since after arriving in a new location in another time zone, the biological clock starts to adjust to the local time, but there is a large degree of variation in the speed and direction at which individuals adapt to new timezones. Amongst other factors, the direction of the time zone crossing can impact the speed of adjustment, with an eastward time zone crossing leading to a slower adjustment than a westward one.

Accurate determination of an individual's circadian rhythm (and its level of disruption) has been proven to be difficult, especially in the operational field. That is why the degree of circadian disturbance is often estimated based on the outcomes of a biomathematical model (which can use work schedules, sleep-wake behaviour, and/or light exposure as inputs) (Skeldon et al., 2017), or is approximated by the local time where the trip began (Gander et al., 2015).

Because it is not feasible for this project to actually measure circadian timing through saliva, urine, blood or core temperature either, we will use a physiology-based mathematical model (Phillips et al., 2011; Phillips & Robinson, 2007) to estimate circadian phase throughout the flight rosters, based on crew members work schedules, sleep-wake times and actual light profiles. The latter will be recorded by actigraphy, using the sophisticated integrated light intensity sensor of the Spectrum Plus ActiWatches. The fact that crew members may wear long sleeves covering the light sensor might pose a challenge. In case light data is considered missing based on auto-correlative analyses, we will check whether data is missing at random vs. non-random, and decide on the best way to handle the missing values, considering imputation (e.g., using auto-regressive models) as well as applying the model with unaltered light input, if amount of missing data is limited (i.e., < 10%).

<u>Chronotype</u>

Inter-individual differences in circadian periods result in different chronotypes, called early, intermediate, and late types. The estimated distribution between these types is 30%, 30%, and 40% respectively (Roenneberg et al., 2019). It appears that early chronotypes are more prone to fatigue due to irregular duty schedules than late or intermediate chronotypes (Jung & Lee, 2015; Storemark et al., 2013; Saksvik et al., 2011). Although it is quite difficult to measure chronotype objectively (e.g. through genetics or hormones), several validated questionnaires exist (Horne & Ostberg, 1976; Shahid et al., 2011). Most of these assume a regular sleep/wake rhythm and do not take irregular work schedules into account. About 10 years ago however, a specific chronotype questionnaire for non-day workers (MCTQshift) has been developed and evaluated (Juda et al., 2013). Another valid possibility to determine one's chronotype is to ask people which chronotype they consider themselves to be, using the self-assessment item from the rMEQ scale (Loureiro & Garcia-Marques, 2015).

In FTL2.0, the evaluation of chronotype has been added to the methodology and will be measured in the baseline questionnaire (Annex B) by means of asking the participants which chronotype they consider themselves to be: an extreme morning-type; a more morning than evening-type; neither a morning nor evening-type; more an evening than morning-type; or an extreme evening-type (Loureiro & Garcia-Marques, 2015). Furthermore, three questions from the MCTQshift will be added to take into account the difference in sleep duration between work days and non-work days (Juda et al., 2013).

2.2 Comparison with methodology used during previous contract

As mentioned above, the experience and lessons learned from the first FTL project have been take into account while developing the measurement approach for the current study (FTL2.0). In Table 2-1 below, the differences between the two studies are listed.

Table 2-1 Comparison op measurement approach between FTL (previous contract) and FTL2.0 (current study).

Variable	Outcome measure	Study			
		FTL		FTL 2.0	
		ActiWatch	Crewalert app	ActiWatch	Fit4Duty app
Fatigue	KSS		x		x
	SP		x		x
Sleep duration, quality and time awake	Sleep duration	x	x	x	x
	Sleep quality		x		x
	Time awake	x	x	x	x
Performance	PVT 5 minute		x		
	PVT 3 minute				x
Workload	VAS (NASA-TLX)				x
	RSME		x		x
Work-related factors	Flight times		x		x
	Sectors		x		x
	TZ crossed		x		x
	Crew composition		x		x

3. Data Collection Procedures

Once the airline has been identified as an eligible candidate for the data collection (Task 1.2 – definition of the target aircrew population), the procedures described in this chapter will be applied based on the decision made above (Chapter 2), the lessons learned from the expert interviews (D1.1) and the previous contract. The following revised data collection procedure, including the presumed timeline, has been composed.

3.1 Procedure at participating airlines

For each selected airline that agrees to participate, a specific, tailored procedure will be determined to suit the FDPs of interest, the targeted population, and the available facilities for that airline. After inclusion of an airline, the Point of Contact (PoC) of the consortium and the designated airline coordinator will decide on the exact timing and duration of the measurement campaign, and the logistical considerations involved. This procedure, however, will be based on the general principles described below, serving as a starting point to work from.

Participant selection and inclusion

The airline coordinator will send out an email with information about the background and objectives of the study, and an explanation of participation to the targeted aircrew, four weeks before the intended start of the measurement campaign. Other possible means of contacting participants if deemed relevant are the airline newsletter or an information leaflet/poster at a convenient location.

The email with the study information will include a link to an online baseline-questionnaire (Annex B This questionnaire includes an informed consent form with clear details on study procedures, the way personal information will be handled (see Chapter 5), and emphasizes the voluntary part of the participation. Participants will not be paid to participate, nor will additional time needed to comply with the data collection procedure count as duty time. The following data will be gathered by means of the baseline questionnaire:

- Contact details (name, date of birth, email address – chapter 5 provides a detailed data handling description);
- Demographics (age, gender, dominant hand);
- Occupational factors (job title, years of experience, employment rate, home base, employer, commuting time);
- Habitual sleep (chronotype, sleep timing on workdays and work-free days, sleep quality);
- Health and lifestyle (general health, work ability, height, weight, caffeine consumption);
- Previous workload (need for recovery, workload during last month); and
- Personal scheduling information during the data collection window (questions tailored per airline).

At the end of the questionnaire, a list of dates for an on-site kick-off meeting will be provided, giving participants the opportunity to indicate at which dates they are available to attend this briefing (the dates will be determined in consultation between the PoC and the airline coordinator). Finally, participants are informed that they will be contacted with further details on their participation.

After two weeks, the baseline questionnaire will be closed and based on the results, the most eligible candidates for the data collection will be invited for the kick-off meeting through an email to their personal email address. Eligibility will be determined based on exposure to the FDPs of interest during the data collection campaign, availability to attend the briefing, and the representativity of the sample (e.g. function, type of crew). The exact date and time(s) for the on-site kick-off meeting will be chosen in coordinance with the airline coordinator and based on the availability of participants. Next to the kick-off meeting information, participants will receive a separate email with their unique participant ID number. It will be stressed that it is vital for participants to store this email because it is their proof that they have filled out their details, are selected for participation, and that the measurement devices necessary for data gathering can be handed out to them (participants without this email/partipant ID will not be receiving devices).

Kick-off meeting

For each airline, the kick-off meeting(s) will be organised in-person at the airline homebase. The goal of this meeting is to instruct participants in using the measurement devices and in following the measurement protocol. The kick-off meeting will be attended by these participants (those who can attend), the airline coordinator or coordinating team, and at least one researcher of the consortium. The measurement equipment that will be distributed to participants will be transported to the airline homebase by the consortium researcher and subsequently stored in a secure location that can only be accessed by the airline coordinator(s).

Each participant will receive:

- A mobile phone with charger (both labelled with a device number);
- An ActiWatch;
- A Quick Reference Card with essential information about the procedure, a QR code to an instruction video, and the contact details of the airline coordinator and the consortium PoC;
- An information booklet with more detailed explanations about the tasks and the timeline; and
- A firm travel case to store and protect all measurement equipment (with a unique case number).

During the kick-off meeting and after being instructed, the participants will receive the travel case with the equipment required for the study. Each travel case has a unique number, which will be linked to the specific participant ID. This way, the researchers will be able to keep track of the status and location of all devices. The file which links the case numbers with the participant IDs will be made accessable online for the airline coordinator, so that the information needed can be diligently updated once the researchers have departed.

It is unlikely that all participants are able to attend the in-person kick-off meeting(s). Hence, personal follow-up meetings with the remaining participants will be organised in which the instructions can be provided by means of the instruction video, and the measurement equipment can distributed by the airline coordinator(s). Subsequently, the coordinator will also be able to answer any potential questions, and register the case and participant IDs in the database.

Data collection period

During the data collection period, participants use the equipment provided to them according to the instructions given during the briefing. They can also review all relevant information using the Quick Reference Card or information booklet that are included in the travel case.

During the data collection period, the airline coordinator and the consortium representative will be in close contact, at least once a week, to evaluate progress on data collection. Incoming data through the app can be continuously monitored through the backend of the app, by the researchers only.

Material collection

After completing their data collection, participants will hand in their measurement equipment at the airline homebase. The airline coordinator will ensure that all returned materials are complete, will update the measurement equipment file, and store the materials in a secure location as agreed upon. In case of an unexpected non-return, the consortium will seek the contact details of the specific participant, and privately inquire about the situation. The complete set of travel cases will preferably be collected by a consortium representative. Devices handed in late may be sent to NLR via FedEx or can be collected at a later point in time. Participants will be personally thanked for their effort by an email to the provided email address.

3.2 Scale and Timeline

Based on the procedures described above, we expect a data collection campaign per specific airline to take 8 to 10 weeks (about 4 weeks of preparation, 4 weeks measurements, 2 weeks closure).

Contrary to data collection in the previous FTL study, in the current project, the participants are targeted specifically to get more relevant and high quality data. By targeting airlines flying certain types of FDPs and the aircrew members flying these types of FDPs, the likelihood of collecting the exact data needed for analysis will increase (D1.2). Also, irrelevant data collection is kept to a minimum, and it is expected to allow for more possibilities to analyse data using within-subjects analyses that have more power (i.e. fewer participants are needed) than between-subjects designs. However, within-subjects analyses cannot handle missing data very well. Therefore, close monitoring of the quantity and quality of data is necessary. Incoming data will be monitored each working day when at least one campaign is running with respect to the amount (dependent on the number and function of participating crewmembers) and quality (e.g. missing values). This procedure, using the backend of the Fit4Duty app, allows for interventions if quality during a data collection campaign with a specific airline proves to be insufficient. For example, if a participant omitted filling out sleep data, this participant can be individually prompted using the app. In case of insufficient data overall, and in coordination with the airline PoC, it could be decided to extend the measurement period and/or recruit additional participants.

If data collection in one FDP nears the required sample size and is in accordance with the quality as requested, the campaign's focus will shift towards the other FDPs. The number of participants that is aimed for per FDP is further described in Chapter 4.

The original project planning, as described in the proposal, anticipated 13 months for the data collection. At the time of writing this deliverable, a three months extension was granted for D1.2 (definition of the target population, to April 1st 2023), for which the data collection phase will start later as well, while the project end-date will remain unchanged (October 1st 2024). However, with the approach described above, we are flexible regarding our FDP-specific data collection efforts, and we expect to have multiple airline-specific data collection campaigns running in parallel. Furthermore, for both the standby FDP and the Controlled Rest study, data can be gathered independent of the main FDPs of interest. Therefore, it will be possible to wrap up the data collection phase in 10-12 months, for which there will be enough time left to analyse the data and to report about the subsequent findings.

4. Study designs

The study designs used to evaluate the effectiveness of the FDPs of interest were revised in comparison with the approach described in the proposal as a result of:

- the definition of baseline (D1.1);
- extensive internal and external (with EASA, the Scientific Committee, and the Mirror Group) discussions of the consortium; and
- the chosen measurement devices and outcome measures.

As is shown in this chapter, each FDP requires a slightly different design, for instance, because data needs to be collected at different times. Another example is that the >13h FDP will benefit from regular measurements during flights. However, this procedure will not be feasible for participants flying a high number of sectors since they will not be able to fill-out multiple questions during (ultra) short flights. This chapter first presents the general study design applied, followed by specificities for each FDP specific study design, and the Controlled Rest study design.

4.1 General study design

The main starting point for the study design is to actively find airlines who regularly fly the FDPs of interest. This process is described in detail in D1.2. In short, a selection was made of airlines that have to adhere to EASA's FTL regulations and are likely to fly at least one of the targeted FDPs. Next, these airlines were approached and asked to participate. If they were willing to participate, they were asked to detail flown flight scheduling data which in turn was analysed to identify if, how often and in which types of flights the target FDPs occur. This information provides the opportunity to select the most eligible airlines, and specific target groups within these airlines that are likely to fly the targeted FDPs regularly.

This procedure however does not allow for a good identification of Controlled Rest and standby metrics in specific airlines. Controlled Rest is not planned and therefore not part of the flight schedule. Standby is part of planned scheduling, but when analysing FDPs in actually flown scheduling data it is not visible which FDPs were preceded by standby duties. Therefore, a specific survey about the prevalence, details, and fatiguing factors of both CR and standby was distributed amongst the crewmembers by the participating airlines to get a better understanding, and to aid in selecting the most eligible airlines for these study objectives. Results of this survey are described in D1.2.

4.1.1 Main research questions

The main goal as described in paragraph 1.1 is to review the effectiveness of FTL regulations by means of an assessment of the impact on aircrew alertness of:

1. Duties of more than 13 hours at the most favourable time of the day, being non-augmented two-pilot duties at the most favorable time of the day (duties that start between 06:00 and 13:29);
2. Duties of more than 11 hours for crew members in an unknown state of acclimatisation, in non-augmented two-pilot duties of 1-2 sectors, without the use of extensions;
3. Duties including a high level of sectors (more than 6) which comply with the maximum daily FDP of duties of 7 or more sectors;
4. An "other-standby" period in combination with an assigned FDP, leading to an extended awake period and/or inadequate sleep; and
5. Controlled Rest. Plus an assessment of the conditions and circumstances under which aircrew members take Controlled Rest.

The overarching goal is to prevent high levels of fatigue and the accompanying safety risks. Since alertness is difficult to assess and fatigue cannot be prevented completely (work always causes fatigue), the regulations will be considered effective if they prevent *high* levels of fatigue. For that purpose a binary categorization of fatigue will be used, following the approach of the previous contract study, classifying scores of ≥ 7 on the KSS as high fatigue.

However, since studies of simulated or real driving find that excessive fatigue leading to a markedly increased accident risk only occurs at KSS 8 or 9 (Ingre et al., 2006; Åkerstedt et al., 2013), results with >7 will be tested as well.

There is less empirical evidence for a cutoff on the SP compared to the KSS. However, airlines have been using values of 5 and above to indicate excessive pilot fatigue, and this cutoff will therefore also be used in the current project (Powell et al., 2011). Furthermore, we also foresee analyses where we may use the full continuous KSS or SP scale to maximize variance (compared to binary approaches) and make it possible to more thoroughly explore associations.

4.1.2 Design

For each FDP type, prevalence of high fatigue will be identified. Next, FDPs that are representative for the FDPs of interest are compared with pre-defined control FDPs.

The objective is to collect complete data from at least 30 flight and 30 cabin crew members for each FDP type. The previous project proved that it might be difficult to obtain sufficient cabin crew, and therefore extra efforts will be made to encourage participation in this group. The foreseen duration of data gathering is two to four weeks and includes both active FDPs and off-duty time. Preferably, each target FDP and its corresponding control condition are measured in the same participant (within subject). However, if this is not achieved, a between subject analysis will be performed. The exact number of FDPs that will be measured depends on the frequency within which these occur during the data collection campaigns at the specific airlines selected. Therefore, the required sample size estimations indicating how many participants are needed to find relevant results are based on both logistic regression model and mixed ANOVA designs (combining the within and between subject approach).

4.1.3 Measurements

As described in chapters 2 and 0, baseline information, such as demographics and habitual sleep, is gathered through an online questionnaire as sent by the airline coordinator. ActiWatches are used for gathering sleep and activity data and all other data is gathered by means of the Fit4Duty app using dedicated smartphones. Figure 4-1 presents an example of the frequency and timing for all measurements when the FDP consists of one flight.

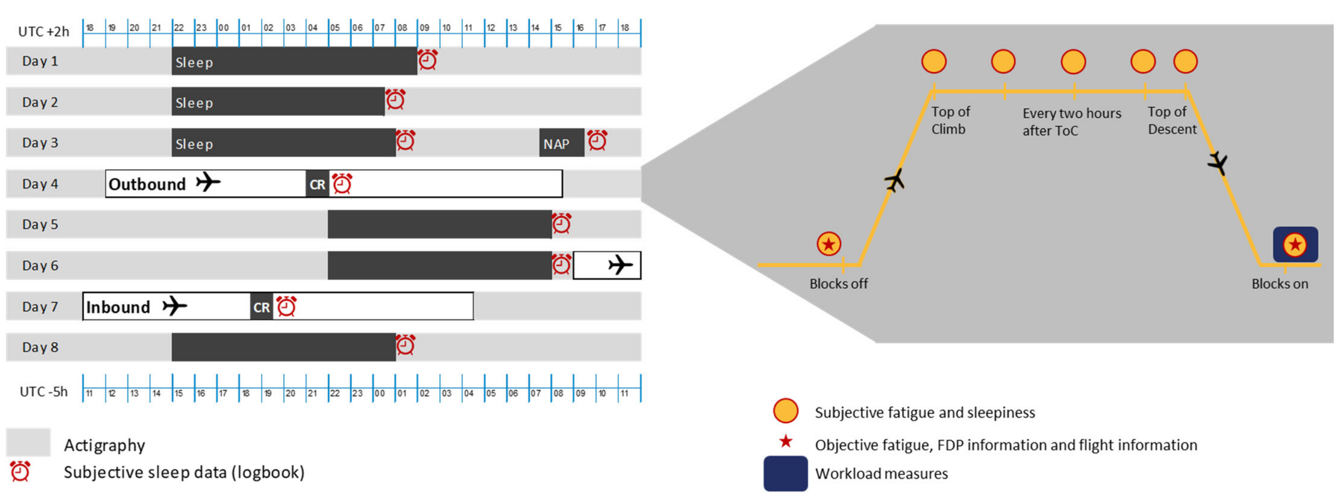

Figure 4-1 Visual example of data collection timing

In this example, day 1 starts 72 hours before the target FDP. Measurements will start after waking up from long sleep by putting on the ActiWatch after the participant has been instructed during the kick-off meeting. As from that moment the ActiWatch will start collecting data and will stop after the participant takes it off at the end of the campaign. This will be 24 hours after the last day of the measurement period.

Every time a participant wakes up from a sleep or napping period, a sleep log will be filled out. Accordingly, after each controlled or in-flight rest period, a set of rest-related questions has to be answered in the app. Data will also be collected before, during and after flights. Before the flight (before blocks off), participants are requested to fill out the available FDP information, perform a PVT and fill out subjective sleepiness and fatigue data. During flights, participants will fill out subjective sleepiness and fatigue around Top of Climb (ToC), every two hours after ToC and around Top of Descent (ToD). Between flights, when the blocks are on, participants fill out subjective sleepiness and fatigue, and they perform the 3-minute PVT. At the end of the last flight of the FDP, participants are asked to rate their subjective sleepiness and fatigue, and to perform the 3-minute PVT again. In addition, they are asked to fill out all relevant flight information, and their experienced workload throughout the FDP.

Please note: this example serves a 'normal' flight and is not applicable for the multiple sectors FDP. For the corresponding study design, we chose a different approach, using other data collection time points (see section 4.4).

4.1.4 Data analysis

Although statistical analysis techniques applied can vary between study designs, the overall aim is to collect target and control FDPs from the same participants. This will allow for statistical analyses based on linear mixed or generalised linear mixed models to achieve a high statistical power while allowing for missing observations and an unbalanced number of observations across participants. If within-subject analysis is not possible, between-subjects analysis of variance and binary logistical regression models will be used as main statistical tests. Potential effect modifiers include travel direction of outbound flight, time awake, time of day, FDP duration, and prior sleep. Random intercepts and potentially random slopes by participant will be included. A-priori sample size estimations are provided using power analyses for each FDP, applying a repeated measures, mixed (between- and within-subject factors) ANOVA design, with parameters: α-level = 0.05, power (1-β) = 0.9, medium effect size of Cohen's f = 0.25 (for mixed ANOVAs).

4.2 Duties of more than 13 hours at the most favourable time of the day

Duties of more than 13 hours at the most favourable time of day are duties that start between 06:00 and 13:29 (Table 2 of ORO.FTL.205) and allow FDPs of up to 13 hours. It is not allowed to plan for longer duties in advance. However, the duration of 13 hours can be extended due to commanders' discretion, split duty or operators' extension (only twice per 7 days). The regulations of this type of duty need to be reviewed for non-augmented crews and cabin crews. Thus, the cockpit crews that are targeted are only two-pilot crews.

The main objective is to investigate whether the current FTL regulations are effective in preventing the occurrence of high fatigue during long FDPs (> 13 hours), that start between 06:00 and 13:29.

4.2.1 Specific research questions

1. What is the frequency of high fatigue during long FDPs (> 13 h) flown at the most favourable time of the day?
2. Is the probability of high fatigue higher during long FDPs (> 13 h) compared to shorter FDPs (≤ 13 h)?
3. What is the relation between high fatigue and time of day in long FDPs (> 13 h)?
4. Does the timing of the FDP affect the difference in the risk of high fatigue between long (> 13 h) and shorter (≤ 13 h) FDPs?

4.2.2 Design

The target FDPs are FDPs longer than 13 hours and the control FDPs those equal to or shorter than 13 hours in duration. The earliest start time is 06:00 and the latest end time 03:29. Figure 4-2 gives an example of posible target and control FDPs that may occur within the period defined as the most favourable time of day for FDPs > 13 h (start time 06:00h - 13:29h).

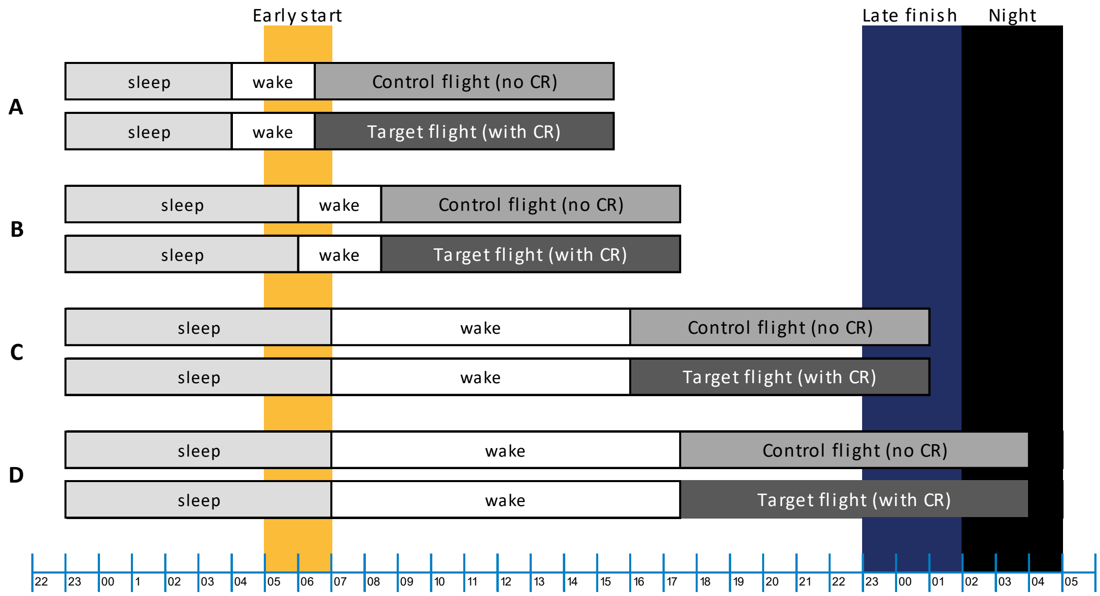

Figure 4-2 Illustrative examples of the target (> 13 h) and control (≤ 13 h) FDPs that may occur within the period of "the most favourable time of the day" set for FDPs > 13 h (i.e., start time 06:00h - 13:29h). The light grey bars indicating sleep periods are estimates of sleep timing and duration in the examples (A-D). Sleep will not be manipulated by the project group any way during the data collection period.

The following situations are represented:

A. **Early start / day finish.** The target FDP starts between 06:00h and 06:59h and has a maximum duration of 13 hours and 15 minutes. The control condition starts in the same period, but the duration is shorter.
B. **Day start / day finish.** The target FDP starts between 07:00h and 08:59h, ends before 23:00h and has a maximum duration of 14 hours. The control FDP starts in the same period and also ends before 23:00h, but is shorter.
C. **Day start / late finish.** The target FDP ends between 23:00h and 01:59h and has a maximum duration of 14 hours. The control condition ends in the same period, but is shorter.
D. **Day start / night finish.** The target FDP ends between 02:00h and 03:29h and has a maximum duration of 14 hours. The control condition ends in the same period, but is shorter.

The aim is to collect four FDPs per participant, see also Table 4-1:

- one earlier (early start/day) and one later (late finish/night) FDP > 13 h (target FDPs); and
- one earlier and one later FDP ≤ 13 h (control FDPs).

The expected number of FDPs per participant will be specify when the duty rosters of the participants are known.

Table 4-1 Number of FDPs aimed for, for analysis of long (> 13h FDPs) and their control FDPs

Number of Measurements	Pilot	Cabin
FDP > 13 hours	30	30
FDP ≤ 13 hours	30	30
Participants	30	30

Design considerations

- If the number of target and control FDPs proves unfeasible to collect from each participant within the timeframe of the study, the FDPs mentioned above will be collected using either an unbalanced within-subject design or a between-subjects design;
- FDPs ending at night are of particular importance because this is when fatigue levels can be expected to be high due to FDP's encroachment on the WOCL (see Figure 4-2); and
- Characteristics of duty rosters other than duration and timing of an individual FDP, such as rest periods and preceding FDPs, will be taken into account in statistical analysis and interpretations of the results.

4.2.3 Proposed data analysis scheme

FDPs and time of day will be considered as categorical variables, with 2 levels for FDP (> 13 hours vs less than 13 hours) and 2 levels for time of day (early/day vs late/night). Fixed effects are 'FDP duration', and 'Time of Day', and the interaction between them. Covariates, when analysing between-subjects, are age, sex, chronotype and other baseline variables. Sample size estimation using power analysis for a 2x2 repeated measures, within-subject ANOVA yielded a total sample size of 46 participants.

4.3 Duties of more than 11 hours for crew members in an unknown state of acclimatisation

Duties of more than 11 hours for crew members in an unknown state of acclimatisation are duties during which a crewmember is not acclimatised to the departure time zone nor to the arrival time zone, thus in an unknown state of acclimatisation. According to ORO.FTL.105(1), crewmembers are in an unknown state of acclimatisation, when 48 hours or more have elapsed since reporting for duty at reference time (i.e., where crew members' duty first started) and when the time difference between reference time and local time (i.e., where crew members start their next duty) is ≥ 4 hours. Table 4-2 gives an overview of the definitions of known and unknown states of acclimatisation, according to ORO.FTL.105(1).

Table 4-2 Definition of known and unknown states of acclimatisation. B = acclimatized to local time of departure time zone. D = acclimatized to local time where crew member starts next duty. X = unknown state of acclimatisation.

Time Difference[1]	Time Elapsed[2]				
	< 48	48 - < 72	72 - < 96	96 - < 120	≥ 120
1 – 3	B	D	D	D	D
4 – 6	B	X	D	D	D
7 – 9	B	X	X	D	D
10 – 12	B	X	X	X	D

The main objective is to investigate the occurrence of high fatigue during duties of more than 11 hours for crew members in an unknown state of acclimatisation. The duration of 11 hours in this FDP is the maximum duration when operating 1-2 sectors. For operators under Fatigue Risk Management (FRM), the maximum daily FDP can be extended to a maximum of 12 hours. These regulations will be studied for non-augmented flight crews and cabin crews.

Jet lag symptoms after westward travel have been shown to be less severe and to dissipate faster than after eastward travel (Herxheimer, 2014; Ambesh et al., 2018). Accordingly, direction of travel could introduce a confounding effect: for instance, if all inbound flights (which, in this study design, are operated in an unknown state of acclimatisation) were to go westward, any effects on fatigue, sleepiness, and cognitive performance might be due to higher jet lag following eastward travel.

4.3.1 Specific research questions

1. What is the frequency of high fatigue during long FDPs (> 11 h) operated in an unknown state of acclimatisation?
2. Is the risk of high fatigue higher during FDPs in an unknown state compared with FDPs in a known state of acclimatisation?
3. Does the risk of high fatigue during unknown-state FDPs depend on duration of the FDP (> 11 h vs. 9-11 h vs. < 9 h)?
4. Does the duration of the FDP (> 11 h vs. 9-11 h vs. < 9 h) affect the (potential) difference in the risk of high fatigue between unknown-state and known-state FDPs?
5. In secondary analyses, we will address the question: Does a known state of acclimatisation (based on mathematical modeling) improve risk predictions of high fatigue for FDPs of different durations (> 11 h vs. 9-11 h vs. < 9 h)?

4.3.2 Design

This FDP has a 2 (state of acclimatisation) X 3 (FDP duration) study design. Levels of 'state of acclimatisation' are 'unknown' versus 'known' and levels of 'FDP duration' are 'shorter than 9 hours', '9-11 hours' or 'longer than 11 hours'.

Data will be assessed from each participant on an outbound flight and the corresponding inbound flight, crossing ≥ 4 time zones and separated by ≥ 48 hours (Figure 4-3). This will result in two FDPs per participant of

---

[1] In hours between reference time and local time where crew member starts next duty.
[2] In hours since reporting at reference time.

similar duration and circumstances (crew, airline, etc.) but in different states of acclimatisation: known-state for the outbound flight and unknown-state for the inbound flight. This approach allows for comparison of the impact of state-of-acclimatisation in a within-subject design. To further examine the effect of FDP duration, flights of different lengths will be included, such that each participant provides data for one category of FDP duration (> 11 h vs. 9-11 h vs. < 9 h). This allows for comparison of the effects of FDP duration in a between-subject design.

This approach will also result in an equal amount of direction of flights (eastward and westward). However, to avoid comparing effects of westward versus eastward flights rather than unknown versus known state of acclimatisation, the travel direction will be balanced as much as possible across outbound flights.

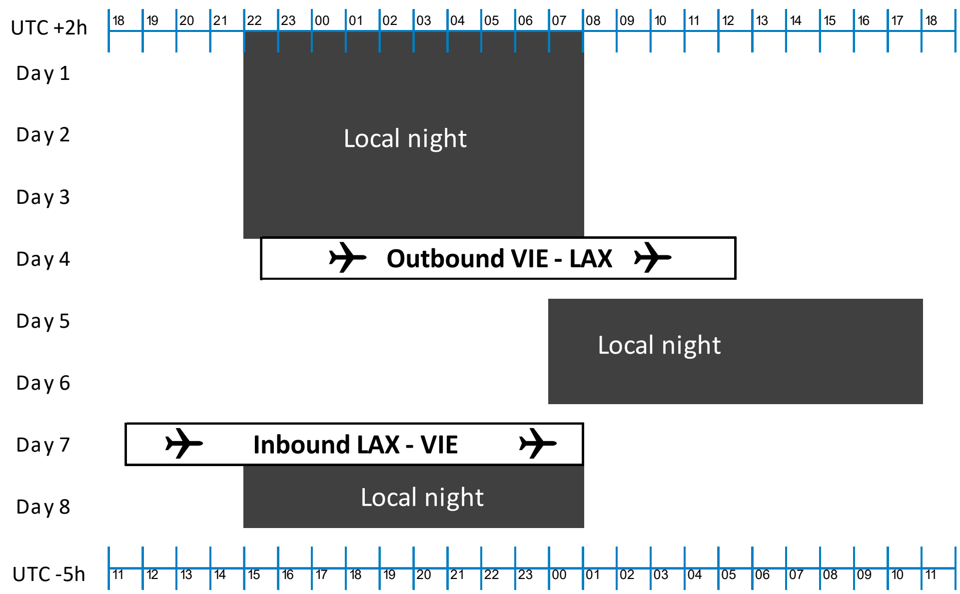

Figure 4-3 For each participant, data will be collected on an outbound flight (operated in a known state of acclimatisation) and the corresponding inbound flight (operated in an unknown state of acclimatisation), with 3 days pre-flight a and 1 day post-flight.

Data collection will include a total of 60 participants (30 pilots and 30 cabin crew members), with 30 participants per level of acclimatization state and 20 participants per level of FDP duration. This results in 60 measured FDPs per state of acclimatisation (known vs. unknown) and 40 measured FDPs per FDP duration (> 11 h vs. 9-11 h vs. < 9 h), see also table Table 4-3.

Table 4-3 Number of FDPs required for analysis of state of acclimatisation and FDP duration

Number of measurements	FDP duration < 9 hours		FDP duration 9 – 11 hours		FDP duration > 11 hours	
	Pilot	Cabin	Pilot	Cabin	Pilot	Cabin
Known state of acclimatisation	10	10	10	10	10	10
Unknown state of acclimatisation	10	10	10	10	10	10
Participants	10	10	10	10	10	10

4.3.3 Data analysis

Primary analyses will treat the two exposures as categorical variables, with 2 levels for state-of-acclimatisation (unknown-state vs. known-state) and 3 levels for FDP-duration (> 11 h vs. 9-11 h vs. < 9 h). Fixed effects include 'state', 'duration', and the interaction 'state x duration'. Covariates (due to the between-subject design) include age, sex, chronotype and other baseline variables. Sample size estimation using power analysis for a mixed ANOVA with a 2-level (acclimatization state, within-subject) x 3-level (FDP duration, between-subject) design yielded a total sample size of 54 participants.

In secondary analyses, we will use a validated mathematical model of circadian and sleep-wake regulation (Skeldon et al., 2017) to estimate circadian phase during the measurement period. The model-estimated circadian phase will then be used to create two additional variables for state-of-acclimatisation. The first variable will be categorical, and use estimated circadian phase to switch participants from an *unknown* state to a *known* state of acclimatisation, that is, whether the participant is acclimatized to reference time (state B, defined as being within ± 1 hour of circadian phase estimated at reference time), acclimatized to local time (state D, defined as being within ± 1 hour of circadian phase estimated at local time), or in a transient state (between states B and D). For instance, if the participant's estimated phase were less than 1 hour away from local time on the day of the inbound-flight FDP, this participant would switch from an unknown state to state D (i.e., acclimatized to local time) in regression analyses. The second variable will be continuous, and use estimated circadian phase to express state of acclimatisation as deviation in hours from reference time. These secondary analyses will allow us to determine the individual degree of circadian misalignment and examine the impact of state-of-acclimatisation at a more in-depth level, with potential insights for individual differences in adaptation to jet lag inducing FDPs.

4.4 Duties including a high level of sectors (more than 6)

These duties are FDPs with high level of sectors (more than 6), that comply with the maximum daily FDP duration, as provided in Table 2 of ORO.FTL.205. This maximum duration depends on the exact number of sectors flown.

The main objective is to investigate the occurrence of high fatigue during this type of FDP. However, very few carriers fly more than 6 sectors with any regularity. Therefore, FDPs with 4, 5 or 6 sectors will also be studied get a better insight into possible relations between a higher number of sectors and fatigue.

4.4.1 Specific research questions

1. What is the prevalence of high fatigue during 1, 2, 3, 4, 5, 6, ≥6 sectors?
2. Is the risk of high fatigue higher during multiple sectors (> 6) compared to fewer sectors (≤ 6)?
3. What is the relation per sector between high fatigue and time of day across sectors (1 to > 6)?
4. Does FDP duty time or time of day modify the above analyses?

4.4.2 Design

The intended design for this type of FDP is a comparison of FDPs with more than 6 sectors with those with less than 6 sectors. The major approach is within-subject (the study mentioned under "Design considerations" below shows that most FDPs allow for within-subject approaches). Information from available carriers indicate that > 6 sectors (or more) almost never occur (see also D1.2). If > 6 sectors are not available, the analyses will focus on mixed model analyses to establish the change in fatigue per sector between 1 and 6 sectors, with extrapolations to > 6 sectors. However, it could turn out that within-subjects analysis is not possible in some carriers, which means that comparisons would then be made between-subjects (not necessarily within the same airline). As a consequence, conditions of FDPs, such as time of day, type of aircraft, occupation and function of participant, will be used to match FDPs found in the target FDP to allow for comparison. Independent of approach, the number of participants aimed for is at least 30 pilots and 30 cabin crew members. Using the power analysis approach indicated above, the analysis arrives at 46 participants for a medium effect size.

Table 4-4 Number of FDPs required for analysis of multiple sectors (>6) and their control FDPs

Number of measurements	Early start / Day	
	Pilot	Cabin
1-6 (possibly 7)	30	30
Participants	**30**	**30**

Design considerations

The data from an older study can be used for re-analysis. This data was collected in 2019 for a report from one carrier to EASA (Åkerstedt et al., 2021). The focus of that specific study was accumulated duty time, but the data that is necessary for the current FTL review, such as number of sectors, duty time per day and sleep hours, were also collected and can therefore be a valuable addition. The dependent variable in this study is the same as that in the rest of the studies in FTL2 (and FTL1), that is KSS (but not SP or PVT).

The study has the advantage that some (24) crew data on 7 sectors were obtained (and a few on 8 and 9 sectors). It also has the avantage that data has been collected at 3 points of day during 7 days and across 4 weeks. Interruptions occur sometimes, due to leave, other duties, etc., but the number of repeated measurements on each individual is considerable and may provide a good source for analyses of the association between number of sectors and fatigue, using a strong intraindividual approach.

The new analysis will contain both binary (< 7 & ≥ 7 on the KSS), as in the published data, but also KSS as a continuous variable.

4.4.3 Measurements

As mentioned in section 4.1.3, the pattern of measurements used for this specific FDP will differ somewhat from the general approach. FDPs with more than 6 sectors will of necessity contain short sectors, and will not contain a cruise phase. Furthermore, much of each sector will be characterized by ascent and descent procedures and are likely to have a high work intensity. It is therefore unlikely that time allows for fatigue ratings to be carried out during flight. Therefore, subjective fatigue ratings are only applied before (blocks off) and after (blocks on) each of the short flights. In addition, the 3-min PVT will be applied before the first flight and after the last flight of the FDP.

4.4.4 Data analysis

A mixed model approach will be used for the analysis, combining within-subject data with between-subject data. The latter may be obtained from different carriers, in which case adjustment for carrier will be applied, in addition to adjustment for FDP duration, time of day and demographics (age, gender, occupation and function). If crew has difficulties rating after blocks off for many sectors, we will apply interpolation to obtain mid-FDP measures. Sample size estimation using power analysis for a mixed ANOVA with a 2-level (number of sectors) x 2-level (FDP timing) design yielded a total sample size of 46 participants.

4.5 Other standby duties

On-call duties such as standby or reserve are duties where the aircrew member is available to receive an assignment for a flight. The main difference between them is that the FDP in a reserve duty is preceded by a rest period of at least 10 hours and the FDP in a standby duty is not. In standby, the crewmember can be at the airport (airport standby), or at another location ("other than airport standby"). When an FDP is not assigned, a rest period follows the standby as specified in ORO.FTL.235.

After consultation with EASA and the Mirror Group, it was decided that this specific study should focus on "other (than airport) standby", and more specifically on the effectiveness of the 18-hour awake rule. According to this rule "the operator's standby procedures are designed to ensure that the combination of standby and FDP do not lead to more than 18 hours awake time" (GM1 CS FTL.1.225(b) Standby)[3]. During "other-standby", crews either stay at home or are accommodated in a suitable accommodation (ORO.FTL.105 (27)). The latter means "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)).

The main aim of the 18-h rule is to ensure an adequate sleep-wake ratio. Over a 24-h period, 18 hours of wakefulness leave 6 hours for sleep, giving a sleep-wake ratio of 0.33. However, the sleep/wake ratio may remain inadequate even when the awake period remains rather short. This is the case when sleep is reduced because of the timing of an FDP assigned during the standby period (e.g., assignment of an early starting FDP). According to CS-FTL1, "Operator procedures for the notification of assigned duties during standby other than airport standby should avoid interference with sleeping patterns if possible" (GM1 CS FTL.1.225(b) Standby)[3].

4.5.1 Main research questions

1. Does the probability of high fatigue significantly increase when the combination of standby and an assigned FPD leads to an extended awake period?
2. Does the probability of high fatigue significantly increase when the combination of standby and an assigned FPD leads to reduced sleep (but not to an extended awake period)?
3. Do the two conditions of an inadequate sleep-wake ratio mentioned above differ from each other in terms of on-duty fatigue?

The first question is our primary question because it directly addresses the issue of an inadequate sleep-wake ratio due to prolonged wakefulness. The second and third questions complement the first one. They shed more light on the inadequate sleep-wake ratio as a factor contributing to on-duty fatigue in the context of "other-standby". The way the data will be collected provides an opportunity to address questions 2 and 3 without significant additional work, as all standby periods within the measurements periods will be recorded (see also "Design considerations" below). The same understandably holds for standby periods not leading to an FDP, as we cannot know in advance if standby leads to an FDP or not.

4.5.2 Design

The target FDPs for standby duties are of two types:

- "other-standbys" that lead to a late finish or night FDP after ***an extended awake period***, see examples A and B in Figure 4-4).

---

[3] https://www.easa.europa.eu/en/downloads/16775/en

- “other-standbys” that lead to a night or early start FDP after reduced sleep, see examples C and D in Figure 4-4).

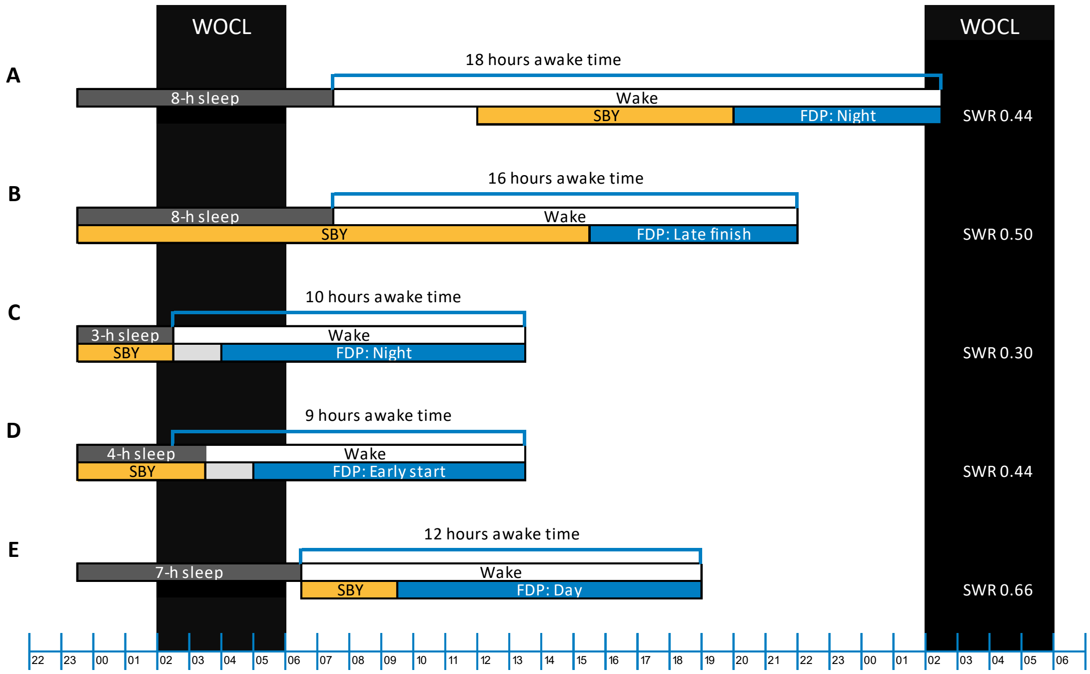

Figure 4-4. Illustrative examples of standbys leading to an FDP. a) A daytime standby followed by a night FDP, b) a night and daytime standby followed by a late finish FDP, c) a night time standby followed by a night FDP, d) a night time standby followed by an e.

“Other-standbys” leading to daytime FDPs will serve as control conditions in terms of on-duty sleepiness and the sleep-wake ratio (see case E in Figure 4-4). Under these conditions neither the amount of sleep is expected to be reduced nor the duration of the awake period to be extended. In addition, the FDP itself will not encroach the WOCL.

The primary aim is to collect as many different kinds of assignments as possible from each participant. These different kinds of assignments are the following:

- night FDP preceded by extended (≈ 18 h) time awake (FDP end time ≥ 02:00h);
- late finish FDP preceded by extended (≈ 16 h) time awake (FDP end time 23:00h-01:59h);
- night FDP preceded by reduced sleep (FDP start time 02:00h-04:59h);
- early start FDP preceded by reduced sleep (FDP start time 05:00h-05:59h); and
- daytime FDP preceded by normal sleep and awake time (FDP starts and ends within the window of 07:00h and 22:59h).

Table 4-5. Number of FDPs aimed for, for the analysis of the other standby FDP, per participant (total n = 60).

Number of measurements	Pilot	Cabin
Night FDP + extended awake	30	30
Late finish FDP + extended awake	30	30
Night FDP + reduced sleep	30	30
Early start FDP + reduced sleep	30	30
Day FDP + normal sleep and awake time	30	30
Participants	30	30

In each type of the target FDPs the sleep-wake ratio (SWR) can be expected to remain rather low. In addition, a part of the assigned FDP may encroach on the WOCL (02:00h – 05:59h), which may further increase on-duty fatigue levels.

Design considerations

"Other-standby duties" not leading to an FDP will also be measured and analysed for explanatory reasons. These data shed light on how much and how well crew members sleep as a function of the duration and timing of the standby period. These results can be used to assess what kind of combinations of standby and FDPs would lead to a lowered sleep-wake ratio (< 0.33) even without violating the 18 h awake time rule.

The target standbys not leading to an FDPs are the following:

- early morning standby (standby starts within the window of 05:00h and 05:59h);
- daytime standby (standby occurs within the window of 06:59h and 22:59h);
- late finish standby (standby ends within the window of 23:00h and 01:59h); and
- night time standby (a part of the standby within the window of 02:00h and 04:59h).

Contrary to the data collection campaigns for the FDPs described above, the data for standby duties will not be collected separately, but derived from the datasets of the other campaigns. If these campaigns do not lead to sufficient data, a separate campaign will be launched to collect more data on "other-standbys" towards the end of the data collection period. The separate campaign will be planned in detail on the basis of the data already collected on "other standbys". Since all airlines who are interested to participate indicated that they schedule other standby for their crewmembers, we think it will be possible to find sufficient candidates for this additional campaign.

4.5.3 Data analysis

The statistical models applied will mostly depend on how many repeated measurements among the participants will be collected. In case sufficiently repeated measurements are found, mainly linear mixed and generalised linear mixed models will be applied for the analysis. In case there is an insufficient number of repeated measurements, between-subjects analysis will be applied instead, by means of variance analysis and binary logistic regression models. In both cases, the conditions with and without an assigned FDP will be analysed separately. Sample size estimation using power analysis for a mixed ANOVA with a 3-level (FDP timing) x 3-level (extended awake/reduced sleep/daytime control) design yielded a total sample size of 45 participants.

4.6 Controlled rest

When unexpected fatigue occurs, at some airlines, a flight crew member can take CR, a period of time during which they are free from flight duties and can use to rest and which may include actual sleep. Crew must follow the airline's controlled rest procedure, which usually states that controlled rest is organised by the commander and may be used if workload permits. Controlled rest taken in this way is not considered to be part of a rest period for purposes of calculating Flight Time Limitations nor used to justify any extension of the duty period.

According to EASA guidance material GM1 CAT.OP.MPA.210, the controlled rest period should be no longer than 45 minutes (in order to limit any actual sleep to approximately 30 minutes) to limit deep sleep and associated long recovery time (sleep inertia). After this 45-minute period there should be a recovery period of 20 minutes to overcome sleep inertia during which control of the aircraft should not be entrusted to the flight crew member.

Controlled rest (CR) is different from in-flight rest: while in-flight rest is planned before a flight with augmented crews and is taken in designated rest facilities (allowing each flight crew member to leave the assigned post to rest and/or sleep). In addition, controlled rest is considered a safety net which is only permitted to be used when crew experience unexpected hgh fatigue. However, in the previous FTL study, a considerable frequency of in-flight naps was observed during the data collection (in 27% of the long night duties), and the assessment of the impact of CR was therefore added as one of the objectives for the current FTL2 study. The subsequent objective is, therefore, to analyse the conditions and circumstances under which CR is taken by flight crew, and to study if it is an effective means to combat unexpected high fatigue. Based on this, it might be possible to recommend CR as a planned fatigue mitigation strategy without implying that the FDP should be extended.

4.6.1 Specific research questions

1. What are the characteristics of flights that have a high prevalence of CR and flights that do not?
2. What are the conditions, circumstances, and reasons for flight crew members to use CR?
3. Is there a difference in high fatigue between flights in which CR is taken and in similar flights in which it is not taken?

4.6.2 Design

Two answer the first two research questions, flights with CR will be analysed to determine the characteristics, conditions, circumstances, and reasons for CR to be used.

In addition, and contrary to the previously described study designs, a comparison will be made between flights instead of FDPs. Flights in which CR is taken are compared to flights during which it is not used. Flight characteristics will be used to create comparable pairs of flights. Flight characteristics that are kept equal are direction, time of day and duration until CR is taken (Figure 4-5).

Results from interviews with airline representatives and aircrews indicate that most people who take CR in a certain type of flight are likely to always do this on that particular type of flight. Therefore, comparisons will be made between subjects.

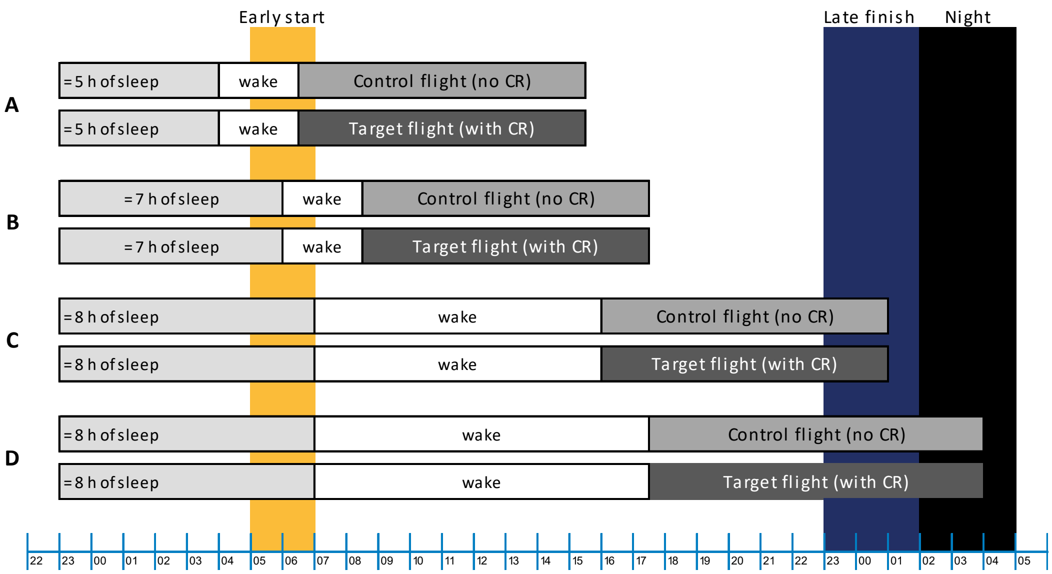

Figure 4-5. Possible Controlled Rest comparisons

Also, the characteristics of flights and demographics that differ between flights and persons taking CR and not taking CR will be compared. Furthermore, an analysis will be done into the level of fatigue after taking CR to determine if CR is effective in preventing high levels of fatigue.

The data collection procedures for the study designs described in the sections above are expected to yield sufficient CR instances in flight crew for analysis. Therefore, targeted recruitment of participants will not be performed. However, if these campaigns do not lead to sufficient data, a separate campaign can be launched to collect more data on CR towards the end of the data collection period. For this, the outcomes of the Controlled Rest and Standby survey (D1.2) can be used to approach the most eligible candidates. Finally, based on the state of the art review (D1.1), the focus of this campaign will be mostly on night flights, since it was found that controlled rest is predominantly taken during these flights.

4.6.3 Data analysis

The characteristics of flights that have a high prevalence of CR, and the conditions, circumstances, and reasons for flight crew members to use Controlled Rest will be described by means of descriptive statistics. To answer the first research question a between-subjects analysis will be applied, by means variance and binary logistic regression models, with sample size esimations using power analysis for between-subject comparison yielding a total sample size of 72 participants. This means that data from at least 36 participating flight crew members using CR and 36 participants not using CR in similar flights are needed for the analysis.

5. Data collection governance scheme

The purpose of the data collection governance scheme is to ensure the quality of the data and to protect natural persons involved. Particular emphasis is placed on the protection of people with regard to the processing of personal data and on the free movement of such data according to Regulation (EU) 2016/679, and on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data according to Regulation (EU) 2018/1725.

The procedures in this section ensure full compliance with the data processing conditions as established by EASA in ANNEX VI (personal data processing records and compliance checklist) of the tender specifications.

The data collection task includes the collection of personal information and experiences and the sharing of anonymous user information with technical partners. Therefore, data processing is carried out according to the principles of Regulation (EU) 2016/679.

- Personal data is processed lawfully, fairly, and in a transparent manner in relation to the data subject ('lawfulness, fairness, and transparency');
- Personal data is adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');
- Personal data is kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed ('storage limitations');
- Personal data is processed in a manner that ensures appropriate security, integrity, and confidentiality of the personal data ('integrity and confidentiality'); and
- Personal data is processed in a manner that enables Regulation (EU) 2016/679 compliance concerning all of these principles ('accountability').

Data will be used solely for the purposes stated. All project data will be stored on a password-protected central database repository at NLR that can be accessed by project members only. The Ethics and Security Manager supervises adherence to procedures for data collection, storage, access, sharing policies protection, retention, and destruction. This is done by ad hoc and scheduled checks on whether and how these procedures are followed. The collected datasets will not be disclosed to third parties during or after the project's lifetime.

5.1 Ethics and Security Manager

The Ethics and Security Manager supports the Project manager in ensuring that the scientific and technical progress of the project complies with the ethical and security standards set out by EU legislation – notably Regulation (EU) 2016/679. The Ethics and Security Manager has the role of Data Protection Officer as described in Article 39 of Regulation (EU) 2016/679, as well as performing additional tasks concerning data processing. As such, the Ethics and Security Manager represent the project teams' accountability of General Data Protection Regulation (GDPR) compliance. More specifically the tasks involve:

- Inform and advise the project members, who carry out data processing, of their obligations pursuant to Regulation (EU) 2016/679 and to other Union or Member State data protection provisions;
- Monitor the ethical and security activities of the project pursuant to Regulation (EU) 2016/679;
- Provide advice where requested as regards the data protection impact assessment and monitor its performance pursuant to Article 35 of Regulation (EU) 2016/679;
- Review potential ethical and security issues which arise as data is being collected;
- Make sure that agreements on the handling of security-related data – if such data are to be used – will be produced (outlined in the PMP) and followed;
- Define the procedures that will be used for participant recruitment (e.g., the process of recruitment, inclusion/exclusion criteria, risks and benefits for the participants). An inclusive approach will be adopted ensuring gender balance for all data sets;

- Assess the whole data collection activity, specifically on any potential effects that may have impacted on the scale and quality of the data collected; and
- Act as the contact point for the supervisory authority on issues relating to processing, including the prior consultation referred to in Article 36 of Regulation (EU) 2016/679, and to consult, where appropriate, with regard to any other matter.

The appointed Ethics and Security Manager within this project is Anneloes Maij (MSc), working at NLR. Dr. Daniel Aeschbach (DLR) fulfilled this role in the previous FTL study, but he will take up a role as an 'external' Ethics and Security advisor this time.

5.2 General procedures

The project data is stored on a separate location on the NLR data server. Access to this part of the server is restricted to NLR project members and NLR ICT personnel. The PM decides which project members will have access and gives instructions to ICT personnel to provide access. The PM maintains a list of project members that have access.

After the formal end of the project the data will remain stored on the NLR data server. The former PM will give instructions to NLR ICT personnel to revoke access to the data by the former NLR project members. NLR personnel can get access to the data after a request to access has been approved by the former PM. Approval will only be given if continued compliance with Regulation (EU) 2016/679 and (EU) 2018/1725 is assured.

The general safety and security measures described in the NLR security plan ("NLR beveiligingsplan"), version 1.3 of 09-04-2021 are applicable.

5.2.1 Data reception and distribution via e-mail

Some of the data, notable flight schedule data, will be received from third parties via e-mail. In these cases, the recipient of the data will copy the data to the project location on the NLR data server. After a check that the data is successfully copied, the original data will be deleted from the NLR mail server.

The project location on the NLR data server is also the vehicle for sharing data across NLR project members. E-mail will not be used to share data across NLR project members, but it is allowed to use e-mail to notify project members on data mutations.

5.2.2 Data storage period

Data will be stored on the NLR data server for a period of 10 years after the formal end of the project. During this period, the former PM will remain responsible for data storage and projection. In case the former PM leaves NLR or obtains another function that requires transfer of responsibilities, the associated department manager will assign the responsibilities for data storage or projection to another NLR employee. At the end of the data storage period the data will be deleted from the NLR data server. No copies will be kept.

5.2.3 Data analysis

Data collected from the various sources will be distributed across the participating research organisations such that each organisation can conduct a relevant part of the data analysis. Only de-identified data will be shared with the participating research organisations. This de-identified data will not contain flight numbers, personnel numbers or any other information that allows identification of individuals. It is the responsibility of the project member that distributes the data to verify that all data is fully de-identified. Before data will be distributed, there will be a check that these data are indeed de-identified. The check will be stored in the project archives.

5.2.4 Data breach procedure

In case a personal data breach is discovered by a project member, he/she will immediately inform the Ethics and Security Manager. The Ethics and Security Manager will subsequently:

- Notify the data breach to the European Data Protection Supervisor;
- Where applicable, communicate the data breach without undue delay to the data subject;
- Notify the NLR security manager if the data breach involves infringement of general NLR data protection measures;
- Document the data breach; and
- Asses the data breach to determine whether additional data protection measures need to be implemented.

5.3 Data to be collected

For the purpose of this study, data of various types are collected and processed.

- Planned and actually flown flight schedules to define the target population;
- Controlled rest and standby questionnaire data;
- ActiWatch data; and
- Data recorded by the Fit4Duty app.

Before participating, data subjects will need to sign an informed consent form which includes transparent information describing how we will process data, the rights subjects have, and contact information of the Ethics and Security Manager. Participants will also be thoroughly instructed on the study purpose, the data collection procedure, and the use of the data collection tools to avoid missing or corrupt data. In all cases the gathered data will be anonymous since the name of the participating individuals is unimportant to the needs of the project.

5.3.1 Planned and achieved rosters to define the target population

Planned and actually flown flight schedules are obtained directly from the airlines via e-mail or a secured filesender server (FileSender.nlr.nl). Once received, the data files are copied to the NLR servers and subsequently immediately removed from the e-mail server. NLR only requests scheduling information from which person names and/or other personal identifying information is removed. In case NLR receives scheduling data that includes names and/or other personal identifying information, this information will immediately be removed. The scheduling information does contain unique identifiers (company provided numbers) for individuals, this will only be used to determine the flight exposure of the individual crew members, and as such the eligibility of the airline to participate in this study.

5.3.2 Controlled rest and standby questionnaire

Data for this survey is collected via the online surveytool LimeSurvey. Limesurvey is a open source frontend application, which runs on the secured NLR servers. Collected data will therefore by no means be accessible for parties other than NLR and/or consortium members with access. Potential participants are invited by the airlines themselves, sharing a general link to the survey, for which no email addresses are known by NLR. After following the link, the first survey question requires the user to give consent to the processing of the data. Gathered data is stored on the NLR server and does not contain any personal identification information.

5.3.3 ActiWatch

ActiWatch data is collected during the data collection campaign, by means of ActiWatches provided by NLR and DLR. These devices will only be worn by individuals who have volunteered to the data collection. Participants can stop using the ActiWatch at any time during the measurement period. The ActiWatch collects wrist movement through an accelerometer, and ambient lighting data through a light sensor, and stores this on an internal storage only. After the ActiWatch has been handed in by the participant, data will be extracted from the device by means of standalone adapter by Respironics™, and directly stored on the NLR servers. The consortium only registers which ActiWatch is provided to which individual to make sure that ActiWatch are returned to the consortium. This registration will be deleted by a project team member immediately after return of the ActiWatch by the individual.

5.3.4 Fit4Duty app

Data is collected on dedicated smartphones via the data collection app Fit4Duty. As per data collection procedures (see 3.1 Procedure at participating airlines and 5.3 Data to be collected), the user of the app is required to give prior consent to the processing of the data, with participants free to stop at any time during the data collection period. Data gathered by the app is directly and securely transferred using SSL- encryption to secure DLR servers, through simcards with a worldwide coverage. NLR has access to the DLR-server through a unique login for the backend of the app. The data gathered will subsequently be copied to the NLR file server. The data collection can be performed anonymously. The consortium only registers which smartphone is provided to which individual to make sure that smartphones are returned to the consortium after the data collection. This registration will be deleted by a project team member immediately after return of the smartphone by the individual.

5.4 Airline cooperation

Within each participating airline, a coordinator (or coordination team, e.g. including representatives of the planning and safety department) of local airline personnel will be assigned to coordinate the data collection. These coordinators will be thoroughly instructed by a project airline PoC assigned within the project team on how to ensure a high-level of quality on the data measured. They will be the main liaison between PoC and participating aircrew members. The instructions will be provided locally at the airline and will emphasize the adherence to the measurement protocol and the reporting and appropriate mitigation of any potential events that might undermine the quality metrics of the data collected. The project team will set up agreements with the participating air operators through which the collaboration between the operator and project team will be defined, particularly in terms of data handling, protection, and exchange with EASA. After completion of the data collection at a specific airline a debriefing will be held with the coordinator/coordination team. The results of the debriefing will act as inputs for the assessment of the whole data collection.

5.5 Ethical approval

Ethical standards and guidelines will be rigorously applied. Ethics approval for the data collection part of the project will be sought through a Dutch Medical Ethical Committee once the data collection procedure as described in this deliverable is approved. It is expected that the procedure, as during the previous FTL project, will be classified as being exempt from further Medical Ethical review.

6. Bibliography

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

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

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

Al-Libawy, H., Al-Ataby, A., Al-Nuaimy, W., & Al-Taee, M. A. (2016). HRV-based operator fatigue analysis and classification using wearable sensors. *In 2016 13th International Multi-Conference on Systems, Signals & Devices (SSD)* (pp. 268-273). IEEE.

Ambesh, P., Shetty, V., Ambesh, S., Gupta, S. S., Kamholz, S., & Wolf, L. (2018). Jet lag: Heuristics and therapeutics. *Journal of Family Medicine and Primary Care, 7(3),* 507.

Arnal, P. J., Thorey, V., Debellemaniere, E., Ballard, M. E., Bou Hernandez, A., Guillot, A., ... & Sauvet, F. (2020). The Dreem Headband compared to polysomnography for electroencephalographic signal acquisition and sleep staging. *Sleep, 43(11),* zsaa097.

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

Arsintescu, L., Chachad, R., Gregory, K. B., Mulligan, J. B., & Flynn-Evans, E. E. (2020). The relationship between workload, performance and fatigue in a short-haul airline. *Chronobiology International, 37(9-10),* 1492-1494.

Barreira, T. V., Kang, M., Caputo, J. L., Farley, R. S., & Renfrow, M. S. (2009). Validation of the Actiheart monitor for the measurement of physical activity. *International Journal of Exercise Science, 2(1),* 7.

Basner M., Mollicone D., & Dinges D. F. (2011). Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronaut;1;69(11-12):*949-959.

Borg, G. (1990). Psychophysical scaling with applications in physical work and the perception of exertion. *Scandinavian Journal of Work Environment & Health;16:*55-58.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews, 44,* 58–75.

Borger, J. N., Huber, R., & Ghosh, A. (2019). Capturing sleep–wake cycles by using day-to-day smartphone touchscreen interactions. *NPJ Digital Medicine, 2(1),* 1-8.

Bostock, S., & Steptoe, A. (2013). Influences of early shift work on the diurnal cortisol rhythm, mood and sleep: within-subject variation in male airline pilots. *Psychoneuroendocrinology, 38(4),* 533-541.

Caldwell Jr, J. A. (1997). Fatigue in the aviation environment: An overview of the causes and effects as well as recommended countermeasures. *Aviation, Space, and Environmental Medicine, 68(10),* 932-938.

Cheung, J., Leary, E. B., Lu, H., Zeitzer, J. M., & Mignot, E. (2020). PSG Validation of minute-to-minute scoring for sleep and wake periods in a consumer wearable device. *PLoS One, 15(9),* e0238464.

Chinoy, E. D., Cuellar, J. A., Jameson, J. T., & Markwald, R. R. (2022). Performance of four commercial wearable sleep-tracking devices tested under unrestricted conditions at home in healthy young adults. *Nature and Science of Sleep, 14,* 493.

Chipchase, S. Y., Lincoln, N. B. and Radford, K. A. (2003) Measuring fatigue in people with multiple sclerosis. *Disability and Rehabilitation, 25 (14).* pp. 778-784. ISSN 0963-8288

Cosgrave, J., Wu, L. J., van den Berg, M., Signal, T. L., & Gander, P. H. (2018). Sleep on long haul layovers and pilot fatigue at the start of the next duty period. *Aerospace Medicine and Human Performance, 89(1),* 19-25.

Dehais, F., Dupres, A., Di Flumeri, G., Verdiere, K., Borghini, G., Babiloni, F., & Roy, R. (2018). Monitoring pilot's cognitive fatigue with engagement features in simulated and actual flight conditions using an hybrid fNIRS-EEG passive BCI. *In 2018 IEEE international conference on systems, man, and cybernetics (SMC)* (pp. 544-549). IEEE.

Depner, C. M., Cheng, P. C., Devine, J. K., Khosla, S., De Zambotti, M., Robillard, R., ... & Drummond, S. P. (2020). Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. *Sleep, 43(2),* zsz254.

Devine, J. K., Chinoy, E. D., Markwald, R. R., Schwartz, L. P., & Hursh, S. R. (2020). Validation of Zulu Watch against Polysomnography and Actigraphy for On-Wrist Sleep-Wake Determination and Sleep-Depth Estimation. *Sensors, 21(1),* 76.

Diaz-Piedra, C., Rieiro, H., Suarez, J., Rios-Tejada, F., Catena, A., & Di Stasi, L. (2016). Fatigue in the military: Towards a fatigue detection test based on the saccadic velocity. *Physiological Measurement, 37(9),* 62–75.

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

Di Stasi, L. L., McCamy, M. B., Martinez-Conde, S., Gayles, E., Hoare, C., Foster, M., . . . Macknik, S. L. (2016). Effects of long and short simulated flights on the saccadic eye movement velocity of aviators. *Physiology & Behavior, 153,* 91–96.

Eldevik, M. F., Flo, E., Moen, B. E., Pallesen, S., & Bjorvatn, B. (2013). Insomnia, Excessive Sleepiness, Excessive Fatigue, Anxiety, Depression and Shift Work Disorder in Nurses Having Less than 11 Hours in-Between Shifts. *PLoS ONE, 8(8),* e70882.

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

Effectiveness of Flight Time Limitation (FTL) – Final Report. MOVE/C2/2016-360. European Commission. (2018).

Fan, J., & Smith, A. P. (2017). The impact of workload and fatigue on performance. In International symposium on human mental workload: Models and applications (pp. 90-105). Springer, Cham.

Ferguson, S. A., Kennaway, D. J., Baker, A., Lamond, N., & Dawson, D. (2012). Sleep and circadian rhythms in mining operators: Limited evidence of adaptation to night shifts. *Applied Ergonomics, 43(4),* 695–701. https://doi.org/10.1016/j.apergo.2011.11.003.

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

Grech, M. R., Neal, A., Yeo, G., Humphreys, M., & Smith, S. (2009). An examination of the relationship between workload and fatigue within and across consecutive days of work: Is the relationship static or dynamic? *Journal of Occupational Health Psychology, 14(3),* 231.

Härmä, M., Laitinen, J., Partinen, M., & Suvanto, S. (1994). The effect of four-day round trip flights over 10 time zones on the circadian variation of salivary melatonin and cortisol in airline flight attendants. *Ergonomics, 37(9),* 1479–1489.

Harris, A., Waage, S., Ursin, H., Hansen, Å. M., Bjorvatn, B., & Eriksen, H. R. (2010). Cortisol, reaction time test and health among offshore shift workers. *Psychoneuroendocrinology, 35(9),* 1339–1347.

Harrison, Y. (2013). The impact of daylight saving time on sleep and related behaviours. *Sleep Medicine Reviews, 17(4),* 285-292.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *In Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, No. 9, pp. 904-908).* Sage CA: Los Angeles, CA: Sage publications.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *In Advances in psychology (Vol. 52, pp. 139-183).* North-Holland.

Herman Hansen, B., Børtnes, I., Hildebrand, M., Holme, I., Kolle, E., & Anderssen, S. A. (2014). Validity of the ActiGraph GT1M during walking and cycling. *Journal of Sports Sciences, 32(6),* 510-516.

Honn, K. A., Satterfield, B. C., McCauley, P., Caldwell, J. L., & Van Dongen, H. P. (2016). Fatiguing effect of multiple take-offs and landings in regional airline operations. *Accident Analysis & Prevention, 86,* 199-208.

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

Hu, X., & Lodewijks, G. (2020). Detecting fatigue in car drivers and aircraft pilots by using non-invasive measures: The value of differentiation of sleepiness and mental fatigue. *Journal of Safety Research, 72,* 173-187.

Ingre, M., Åkerstedt, T., Peters, B., Anund, A., Kecklund, G., & Pickles, A. (2006). Subjective sleepiness and accident risk avoiding the ecological fallacy. *Journal of Sleep Research, 15(2),* 142-148.

Juda, M., Vetter, C., & Roenneberg, T. (2013). The Munich ChronoType Questionnaire for Shift-Workers (MCTQShift). *Journal of Biological Rhythms, 28(2),* 130–140.

Jung, H. S., & Lee, B. (2015). Contributors to shift work tolerance in South Korean nurses working rotating shift. *Applied Nursing Research, 28(2),* 150–155.

Lahtinen, T. M. M., Koskelo, J. P., Laitinen, T., & Leino T. K. (2007). Heart rate and performance during combat missions in a flight simulator. *Aviation, Space & Environmental Medicine, 78(4):*387-391

LeDuc, P. A., Greig, J. L., & Dumond, S. L. (2005). Involuntary eye responses as measures of fatigue in U.S. Army apache aviators. *Aviation Space & Environmental Medicine, 76(7, Suppl.),* C86–C91.

Lee, D. H., Jeong, J. H., Kim, K., Yu, B. W., & Lee, S. W. (2020). Continuous EEG decoding of pilots' mental states using multiple feature block-based convolutional neural network. *IEEE access*, 8, 121929-121941.

Leggatt, A. (2005). Validation of the ISA (Instantaneous Self Assessment) subjective workload tool. *In Contemporary Ergonomics 2005.* Taylor & Francis.

Li, F.; Chen, C.-H.; Xu, G.; Khoo, L. P.; & Liu, Y. (2019). Proactive mental fatigue detection of traffic control operators using bagged trees and gaze-bin analysis. *Advanced Engineering Informatics, 42,* 100987.

Loh, S., Lamond, N., Dorrian, J., Roach, G., & Dawson D. (2004). The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments, & Computers, 36 (2):* 339-346.

Loureiro, F., & Garcia-Marques, T. (2015). Morning or evening person? Which type are you? Self-assessment of chronotype. *Personality and Individual Differences, 86,* 168–171.

Lowndes, B. R., Forsyth, K. L., Blocker, R. C., Dean, P. G., Truty, M. J., Heller, S. F., ... & Nelson, H. (2020). NASA-TLX assessment of surgeon workload variation across specialties. *Annals of surgery, 271(4),* 686-692.

Marino, M., Li, Y., Rueschman, M. N., Winkelman, J. W., Ellenbogen, J. M., Solet, J. M., ... & Buxton, O. M. (2013). Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep, 36(11),* 1747-1755.

McDevitt, B., Moore, L., Akhtar, N., Connolly, J., Doherty, R., & Scott, W. (2021). Validity of a Novel Research-Grade Physical Activity and Sleep Monitor for Continuous Remote Patient Monitoring. *Sensors, 21(6),* 2034.

McNames, J. and M. Aboy (2006). Reliability and accuracy of heart rate variability metrics versus ecg segment duration. *Medical and Biological Engineering and Computing, 44(9),* 747–756.

McKinley, R. A., McIntire, L. K., Schmidt, R., Repperger, D. W., & Caldwell, J. A. (2011). Evaluation of eye metrics as a detector of fatigue. *Human Factors, 53(4),* 403–414.

Montgomery-Downs, H. E., Insana, S. P., & Bond, J. A. (2012). Movement toward a novel activity monitoring device. *Sleep and Breathing, 16,* 913-917.

Natvik, S., Bjorvatn, B., Moen, B. E., Magerøy, N., Sivertsen, B., & Pallesen, S. (2011). Personality factors related to shift work tolerance in two-and three-shift workers. *Applied Ergonomics, 42(5),* 719-724.

Neu, D., Mairesse, O., Montana, X., Gilson, M., Corazza, F., Lefevre, N., ... & Verbanck, P. (2014). Dimensions of pure chronic fatigue: psychophysical, cognitive and biological correlates in the chronic fatigue syndrome. *European Journal of Applied Physiology, 114(9),* 1841-1851.

Nguyen, T., Ahn, S., Jang, H., Jun, S. C., & Kim, J. G. (2017). Utilization of a combined EEG/NIRS system to predict driver drowsiness. *Scientific Reports, 7(1),* 1-10.

Patke, A., Murphy, P. J., Onat, O. E., Krieger, A. C., Özçelik, T., Campbell, S. S., & Young, M. W. (2017). Mutation of the Human Circadian Clock Gene CRY1 in Familial Delayed Sleep Phase Disorder. *Cell, 169(2),* 203-215.e13.

Peißl, S., Wickens, C. D., & Baruah, R. (2018). Eye-tracking measures in aviation: A selective literature review. *The International Journal of Aerospace Psychology, 28(3-4),* 98-112.

Petrilli, R. M., Roach, G. D., Dawson, D., & Lamond, N. (2006). The sleep, subjective fatigue, and sustained attention of commercial airline pilots during an international pattern. *Chronobiology International, 23(6),* 1357-1362.

Phillips, A. J., Czeisler, C. A., & Klerman, E. B. (2011). Revisiting spontaneous internal desynchrony using a quantitative model of sleep physiology. *Journal of Biological Rhythms, 26(5),* 441–453.

Phillips, A. J. K., & Robinson, P. A. (2007). A quantitative model of sleep-wake dynamics based on the physiology of the brainstem ascending arousal system. *Journal of Biological Rhythms, 22(2),* 167–179.

Powell, D., Spencer, M. B., & Petrie, K. J. (2011). Automated collection of fatigue ratings at the top of descent: A practical commercial airline tool. *Aviation, Space, and Environmental Medicine, 82(11),* 1037–1041.

Powell, D., Spencer, M. B., Holland, D., Broadbent, E., & Petrie, K. J. (2007). Pilot fatigue in short-haul operations: Effects of number of sectors, duty length, and time of day. *Aviation, Space, and Environmental Medicine, 78(7),* 698-701.

Powell, D., 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.

Powell, D., Spencer, M. B., & Petrie, K. J. (2010). Fatigue in airline pilots after an additional day's layover period. *Aviation, Space, and Environmental Medicine, 81(11),* 1013-1017.

REGULATION (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016) *Official Journal of the European Union L119*, p. 36.

Reis, C., Mestre, C., Canhão, H., Gradwell, D., & Paiva, T. (2016). Sleep complaints and fatigue of airline pilots. *Sleep Science, 9(2),* 73–77.

Roach, G. D., Petrilli, R. M., 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.

Roenneberg, T., Pilz, L. K., Zerbini, G., & Winnebeck, E. C. (2019). Chronotype and Social Jetlag: A (Self-) Critical Review. *Biology, 8(3), Article 3.*

Rowland, L. M., Thomas, M. L., Thorne, D. R., Sing, H. C., Krichmar, J. L., Davis, H. Q., . . . Belenky, G. (2005). Oculomotor responses during partial and total sleep deprivation. *Aviation, Space, and Environmental Medicine, 76 (7),* 104–113.

Said, S., Gozdzik, M., Roche, T. R., Braun, J., Rössler, J., Kaserer, A., ... & Tscholl, D. W. (2020). Validation of the raw National Aeronautics and Space Administration Task Load Index (NASA-TLX) questionnaire to assess perceived workload in patient monitoring tasks: pooled analysis study using mixed models. *Journal of Medical Internet Research, 22(9),* e19472.

Saksvik, I. B., Bjorvatn, B., Hetland, H., Sandal, G. M., & Pallesen, S. (2011). Individual differences in tolerance to shift work–a systematic review. *Sleep Medicine Reviews, 15(4),* 221-235.

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

Samel, A., Wegmann, H. M., Vejvoda, M., Drescher, J., Gundel, A., Manzey, D., & Wenzel, J. (1997). Two-crew operations: stress and fatigue during long-haul night flights. Aviation, Space, and Environmental Medicine, 68(8), 679-687.Samel, A., & Wegmann, H. M. (1987). Desynchronization and internal dissociation in aircrew. *Ergonomics, 30(9),* 1395-1404.

Samel, A., Wegmann, H. M., & Vejoda, M. (1995). Jet lag and sleepiness in aircrew. *Journal of Sleep Research, 4*:30-36.

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

Sauvet, F., Bougard, C., Coroenne, M., Lely, L., Van Beers, P., Elbaz, M., Guillard, M., Leger, D., & Chennaoui, M. (2014). In-flight automatic detection of vigilance states using a single EEG channel. *IEEE transactions on bio-medical engineering, 61(12),* 2840–2847.

Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2011). Karolinska sleepiness scale (KSS). *In STOP, THAT and one hundred other sleep scales (pp. 209-210).* Springer, New York, NY.

Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2011). Munich chronotype questionnaire (MCTQ). *In STOP, THAT and One Hundred Other Sleep Scales (pp. 245-247).* Springer, New York, NY.

Shariat, A., Cleland, J. A., Danaee, M., Alizadeh, R., Sangelaji, B., Kargarfard, M., ... & Tamrin, S. B. M. (2018). Borg CR-10 scale as a new approach to monitoring office exercise training. *Work, 60(4),* 549-554.

Shin, M., Swan, P., & Chow, C. M. (2015). The validity of Actiwatch2 and SenseWear armband compared against polysomnography at different ambient temperature conditions. *Sleep Science, 8(1),* 9-15.

Skeldon, A. C., Phillips, A. J., & Dijk, D.-J. (2017). The effects of self-selected light-dark cycles and social constraints on human sleep and circadian timing: A modeling approach. *Scientific Reports, 7,* 45158.

Storemark, S. S., Fossum, I. N., Bjorvatn, B., Moen, B. E., Flo, E., & Pallesen, S. (2013). Personality factors predict sleep-related shift work tolerance in different shifts at 2-year follow-up: a prospective study. *BMJ Open, 3(11),* e003696.

Tresguerres, J. A., Ariznavarreta, C., Granados, B., Martín, M., Villanúa, M. A., Golombek, D. A., & Cardinali, D. P. (2001). Circadian urinary 6-sulphatoxymelatonin, cortisol excretion and locomotor activity in airline pilots during transmeridian flights. *Journal of Pineal Research, 31(1),* 16-22.

Van den Berg, M. J., Signal, T. L., & Gander, P. H. (2019). Perceived workload is associated with cabin crew fatigue on ultra-long range flights. *The International Journal of Aerospace Psychology, 29(3-4),* 74-85.

Van den Berg, M. J., Signal, T. L., Mulrine, H. M., Smith, A. A., 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.

Van Drongelen, A., Boot, C. R., Hlobil, H., Van Der Beek, A. J., & Smid, T. (2017). Cumulative exposure to shift work and sickness absence: associations in a five-year historic cohort. *BMC Public Health, 17(1),* 1-12.

Zhang Y, Kotejoshyer R, Punnett L, Buchholz B (2018). Physical workload, leisure-time physical activity, musculoskeletal disorders, and sleep quality among rehabilitation employees. *Sleep;41(1),* A341.

Annex A Review on methods to measure fatigue, sleep, workload, and related factors in the field.

Fatigue

There are many different methods to measure fatigue, which can be roughly divided into self-assessment scales, eye-tracking, and (psycho)physiological measurement techniques.

Self assessment scales

The two self-assessment scales, which are most often used in fatigue research, are the Karolinska Sleepiness Scale or KSS (Åkerstedt & Gillberg, 1990) and the Samn-Perelli (SP) crew status check (Samn & Perelli, 1982). The operators can use these scales to indicate their experienced fatigue/sleepiness level on a 9-point, respectively, a 7-point scale. The KSS scale ranges from “extremely alert” to “very sleepy, great effort to keep alert, fighting sleep”. The SP ranges from and “fully alert, wide awake” to “completely exhausted, unable to function effectively”. The scales are quick research tools to measure fatigue and have, been regularly used in fatigue research (e.g. Åkerstedt et al., 2014; Cosgrave et al., 2018), including studies in aviation (e.g. Åkerstedt et al., 2021; Honn et al., 2016; Petrilli et al., 2006; Samel et al., 1997). Experts indicate that both scales are highly recommended, with a slight preference of the KSS over the SP scale due to the higher granularity of the KSS scores. Interesting to add is that level 7 of the KSS (sleepy – but no difficulty remaining awake) indicates the start of electroencephalographic (EEG) and electrooculographic (EOG) changes representing sleepiness, while levels 8 and 9 are associated with high probability of line crossings on real roads and accidents in simulators. This makes it also possible to use this scale for a binary categorization of fatigue as well (classifying scores of ≥ 7 as high fatigue and scores of < 7 as low fatigue), as addition to using it as a continuous variable. Both the KSS and SP have the advantage of being easy to use in an operational environment and correlate well with decisions by crew to report fatigue.

Eye-tracking, and psychophysiological measurement techniques

Eye-tracking and other (psycho)physiological measurement techniques can provide valuable insight into the level of fatigue experienced by aircrew (Peißl et al., 2018).

According to a review study by Hu & Lodewijks (2020), electroencephalography (EEG) has been widely acknowledged as the ‘‘gold standard” to assess fatigue. EEG is a technique that measures the electrical activity of the brain. Its recordings reflect the sum of electrical activity from large populations of neurons in the brain, and are typically displayed as waveforms, with different frequencies of electrical activity (such as alpha and theta powerbands) corresponding to different types of brain activity. Various studies demonstrated that spectral power in alpha and theta band increases as a person feel fatigued. For example, Sauvet et al. (2014) used EEG to monitor pilot fatigue. The EEG recordings of pilots were analyzed during a long-haul flight. They conducted a 10-h long-haul EEG-based experiment during real flight in which they proved the ratio of various powerbands to be a fatigue indicator for pilots.

Eye-tracking measures the saccadic movement and gaze patterns of the human eyes and can provide information on changes in cognitive processing speed and accuracy as fatigue sets in. Changes in eye movements, such as decreased saccadic eye movements (Rowland et al., 2005; LeDuc et al., 2005; Di Stasi et al., 2016; Diaz-Piedra et al., 2016), increased blink rate (Li et al., 2019), increased fixations (McKinley, 2011), and/or increased total eye closure duration (McKinley, 2011), can be used to detect the onset of fatigue and drowsiness. An alternative method to measure eye movement is by means of electrooculography (EOG), which is a method of measuring eye movements by detecting changes in the electrical potential between two electrodes placed on the skin near the eyes. It provides less information about the eye movements compared to eye-tracking (such as location of eye fixations) but it is possible to detect, for example, blink rate (Borghini et al., 2014).

Functional near-infrared spectroscopy (fNIRS) is a neuroimaging technique that can also be used to detect fatigue in aircrew. fNIRS measures changes in blood oxygeenation levels in the brain, which can be associated with changes in cognitive and physiological processes indicating fatigue and drowsiness. In a study by Dehais et al. (2018) fNIRS was used in combination with EEG, and pilots were monitored while performing a simulated flight task. The results showed that the oxygenation levels in certain regions of the brain, such as the prefrontal cortex, decreased as the pilots became more fatigued.

An electrocardiogram (ECG) can also be used as indicator for fatigue. It's a technique that measures the electrical activity of the heart by placing electrodes on the body. The electrodes are connected to a device, which amplifies and records the electrical signals. In particular, it has been described that heart rate (HR) and heart rate variability (HRV) are related to fatigue. HRV measures the variation in the time between heartbeats and can provide information on the level of stress and arousal. If HRV decreases the subject experiences more high-frequency than low-frequency heartbeats – and this can be an indication of the emergence of fatigue (Al-Libawy et al., 2016). A measurement period of at least five minutes is advised to calculate someone's HRV (McNames & Aboy, 2006), making it difficult to measure very specific moments in time.

Performance as an indicator of fatigue

A worse than usual performance can be an indicator of fatigue (e.g. Caldwell, 1997). Different methods exist to measure performance degradation, obviously depending on the task being performed. For aviation field studies, the following non-exhaustive methods listedcould be considered.

The psychomotor vigilance task (PVT) is a validated test to measure sustained attention or vigilance in an objective manner (Dinges & Powell, 1985; Loh et al. 2004). For up to 10 minutes, the subjects need to respond as fast as possible to a visual stimuli. Based on the subjects' reaction times and lapses, an indication of their level of effectiveness (being the opposite) and thus vigilance can be made. The PVT is a common technique to measure vigilance in civil and military pilots (e.g. Petrilli et al., 2006; Roach et al., 2012; Van den Berg et al., 2015; Honn et al., 2016). More recently, shorter versions of the PVT have been developed and validated which might be more suitable to use when studying aircrew on duty. The compliance with the five-minute PVT proved to be quite low in the previous FTL project. A more suitable option therefore could be the even shorter three-minute version of the PVT (Basner et al., 2011; Elmenhorst et al., 2012), as long as participants are well instructed when and how to perform this task during duties.

Sleep

When studying the effect of the current flight time limitations on fatigue, it is important to take sleep length and quality into account since sleep has a direct influence on fatigue levels (e.g. Eldevik et al., 2016). Furthermore, Flight Time Limitations and therefore aircrew schedules are partly based on the recuperative effect of sleep, and the maximum amount of sleep possible in between FDPs. However, the actual amount of sleep someone experiences can be disrupted by various factors (such as an inadequate sleep environment). It is therefore important to also measure the actual amount, quality and timing of the pilots' sleep. Several measurement techniques exist to measure a person's sleep duration and quality.

Polysomnography (PSG)

Polysomnography (PSG) has been widely acknowledged as the ''gold standard" to assess sleep (duration and quality). PSG measures various physiological parameters while the person is sleeping, including brain activity, eye movements, muscle activity, heart rate, and respiration. However, despite the potential benefits of these instruments (PSG and its various physiological parameter), they are not practical for use in a large scale study. The use requires specialized equipment and trained technicians, is expensive and the data collection process can be time-consuming and intrusive.

Some Commercial Of The Shelf (COTS) devices can also assess sleep (quality and duration) based on various (psycho)physiological parameters such as brain activity. These devices are not yet widely used for scientific research, although validity studies show promising results in accuracy, sensitivity and specificity, compared to

PSG. For example, Dreem 2 is a headband capable of measuring powerbands using EEG sensors. In addition, it measures heart rate and movement, position and breathing frequency using a 3D accelerometer. According to a study of Arnal et al. (2020), Dreem 2 reached similar performance compared to PSG data, only with a slight underestimation of wake time. However, these devices a relatively new to the market and more research is needed to provide more evidence on its validity and reliability.

Actigraphy

Sleep duration and time awake can be estimated using actigraphy. Actigraphy involves measuring the physical activity of an individual over several days or weeks, typically using a wrist-worn actigraphy device with an accelerometer. The accelerometer provides a motion signal from which reliable wake and sleep periods can be determined. Some actigraph devices such as the ActiWatch series (Philips Respironics, Inc.; Murrysville, PA, USA) can also record light intensity in lux via an integrated sensor which supports the identification of the wake and sleep periods. The ActiWatch series has been validated and widely used for detection of sleep duration and sleep quality in aviation (Shin, Swan & Chow, 2015; Devine et al., 2020). According to the study of Cheung et al. (2020), ActiWatches have an accuracy of 88.7%±4.5%, sleep sensitivity of 92.6%±5.2%, and sleep specificity of 60.5%±20.2%, compared to polysomnography (PSG). Actigraphy has a relatively low specificity for differentiating sleep from motionless wakefulness, resulting in a possible overestimation of total sleep time (TST) and underestimation of wake after sleep onset (WASO) time (Marino et al., 2013; Montgomery-Downs et al., 2012).

In addition, wearable devices such as Fitbit or the Oura ring can also record heart rate (variability) metrics, which, besides measuring sleep length, support in assessing the sleep quality based on the different durations of various sleep stages. Chinoy et al (2022) compared ActiWatches with four commercial wearables and found that wearables such as Fitbit and Oura ring compared favorably to actigraphy in wake detection. However, high variability in sleep stage-tracking performance suggested that wearables are still best utilized for tracking sleep-wake outcomes and not sleep stages.

While a lot of commercial wearables exhibit promising performance for tracking sleep-wake in real-world conditions, this supports the consideration of using these devices as an alternative to actigraphy in scientific studies. Nevertheless, a group of sleep experts concluded in a workshop that wearables need to be further validated, and that accelerometry-based sleep trackers can be preferred if the main interest involves sleep/wake detection, availability of raw data, data privacy, and the battery life (Depner et al., 2020).

Sleep log

Another technique with which sleep-wake characteristics can be measured is through self-report in a sleep diary. A sleep diary or sleep log is a record of an individual's estimates of sleep and wake times, typically made over a period of several weeks. It is an inexpensive technique which provides a subjective estimate of sleep quantity and/or quality. In most applied research, actigraphy and sleep diary data are collected in parallel. The accuracy of self-reported sleep data appears to be acceptable – as was confirmed during the previous FTL study, where a comparison between the sleep logs and actigraphy recordings was made and a strong positive correlation was found ($r = .888$, $p < .001$). Due to its non-intrusiveness and the option to fill it out at a convenient moment, sleep logs have often been used in studies on cabin crew and airline pilots on duty (Åkerstedt et al., 2021; Petrilli et al., 2006; Roach et al., 2012; Samel et al., 1997; Van den Berg et al., 2015; Van den Berg et al, 2019).

Tappigraphy

In a study by Borger et al. (2019), a comparison was made between two standard methods (i.e. actigraphy and a sleep log) and a tappigraphy-based algorithm. Tappigraphy is the analysis of tapping, scrolling and other interactions with a smartphone. The researchers found strong correlations between the sleep times determined by actigraphy versus the gaps in phone usage found by the tappigraphy algorithm. However, more research about the validity and practical feasibility of this technique in the field is needed.

Circadian rhythm

It is well established that the effects of time zone changes on circadian processes are significant after the crossing of three or more time zones (Samel et al., 1995). It has also become clear however that imposing a one hour time difference (e.g. through the introduction of day light saving time) can lead to sleep disruption and a change in health and regulatory behaviours (Harrison, 2013). This is relevant for aircrew since after arriving in a new location in another time zone, the biological clock starts to adjust to the local time, but there is a large degree of variation in the speed and direction at which individuals adapt to new time zones. Amongst other the factors, the direction of the time zone crossing can impact the speed of adjustment, with an eastward time zone crossing leading to a slower adjustment than a westward one. When we cross time zones and our biological clock is acclimating to the new time zone we experience jet lag, which is characterised by sleep disturbances, fatigue and performance decrements) (Berger & Hobbs, 2006). Information on the level circadian disruption is thus important to include in the data collection, particularly addressing the FDP looking into the state of acclimatisation.

Hormones

A more direct possibility of determining circadian disturbance is by means of evening melatonin levels in dim light. The circadian rhythm of melatonin in saliva (or plasma) is a prominent output of the suprachiasmatic nucleus (SCN) function, the endogenous circadian pacemaker (the biological clock). Melatonin levels and/or Dim Light Melatonin Onset (DLMO - a measure of timing of the circadian clock) have been used in a variety of industrial settings to measure circadian disruption as a result of shift schedules and occupational exposures (Ferguson et al., 2012; Harris et al., 2010). In order to determine DLMO, saliva, blood or urine has to be collected daily . Studies that have applied this method with flight crew members in real life are scarce (Härmä et al., 1994).
Another clock-regulated hormone, cortisol, typically follows an inverse U-shaped rhythm with low levels during the night and higher levels during the day (Bostock & Steptoe, 2013). Cortisol can be measured in urine, blood and saliva, and has been used in some aircrew studies before (Bostock & Steptoe, 2013; Tresguerres et al., 2001).
Both melatonin and cortisol measurements are costly, require strict protocols and are intrusive for the participants for which they are often not feasible in complex operational environments, especially when data collection has to be performed unsupervised (such as in the current study).

Core body temperature

Similar to hormonal levels, core body temperature fluctuates in a circadian rhythm and is a reliable indicator of circadian timing (Samel and Wegmann; 1987). The circadian rhythm in core body temperature is endogenous and occurs independent of sleep. It can be measured continuously with a temperature sensor, for instance by means of one being placed on the chest, or being swallowed.

Accurate determination of an individual's circadian rhythm (and its level of disruption) has been proven to be difficult, especially in the operational field. That is why the degree of circadian disturbance is often estimated based on the outcomes of a biomathematical model (which can use work schedules, sleep-wake behaviour, and/or light exposure as inputs) (Skeldon et al., 2017), or is approximated by the local time where the trip began (Gander et al., 2015).

Workload

Both low and high workload can lead to a faster onset of fatigue (Fan & Smith, 2017; Grech et al., 2009). Subjective rating scales to measure workload are relatively cost-effective and easy to apply. However, the validity and reliability of some subjective measures can vary. The three subjective measurements techniques below have proven to be valid and are commonly used to measure workload. Actigraphy will also be considered, as an objective technique for workload measurement..

NASA Task Load Index (TLX)

This rating scale was developed by NASA (Hart & Staveland, 1988; Hart, 2006) and consists of six component scales: ‘mental demand’, ‘physical demand’, ‘temporal demand’, ‘performance’, ‘effort’ and ‘frustration’. An average of these six scales, weighted to reflect the contribution of each factor to the workload of a specific activity from the perspective of the rater, is proposed as an integrated measure of overall workload. A Visual Analog Scale (VAS) is used to rate each component, but numerical values are not displayed. Values range from 0 = “Low” to 100 = “High” (except for the inverted performance scale). Although it can take some time to fill in the six scales, the NASA TLX has often been used in various high-risk professional environment studies such as, but not limited to, the medical sector (Lowndes et al., 2020; Said et al., 2020) and aviation sector (Arsintescu et al., 2020;Samel et al., 1997).

Borg CR-10

The Borg CR-10 is another subjective method, specifically for measuring physical workload (Borg, 1990; Shariat et al., 2018). Participants can use this scale to indicate their experienced physical workload on a 10-point scale, ranging from “nothing at all” to “extremely strong”. A high correlation was found between the NASA TLX and the Borg CR-10 (Lahtinen, et al., 2007), and the VAS and the Borg CR-10 (Borg, 1990).

Instantaneous Self-Assessment (ISA)

The ISA provides a simple unambiguous workload measure throughout a task with minimal intrusion (Leggatt, 2005). It consists of a simple 5-point scale reaching from 1 (=Underutilised) to 5 (=Excessive), represented by five buttons. The rater responds to prompts at predefined intervals by pressing one of these buttons. A higher average score on the ISA indicates higher workload. Because responses are gathered in real-time, ISA ratings are not subject to memory decay and bias that can occur with measures that rely on post-task assessments.

Rating Scale Mental Effort (RSME)

The RSME is anoter subjective rating scale on mental effort (Zijlstra, 1993). The bar runs from ‘absolutely no effort’ to ‘extreme effort’ (see Figure 2-1). If used for aircrew, the level of mental effort experienced in a particular duty can be analysed for mean and peak levels. A peak RSME rating of above 100 is classified as high workload; an average rating of below 30 as low workload. The RSME can be administered using a paper logbook or in a mobile data collection app, which is in accordance with the previous FTL study.

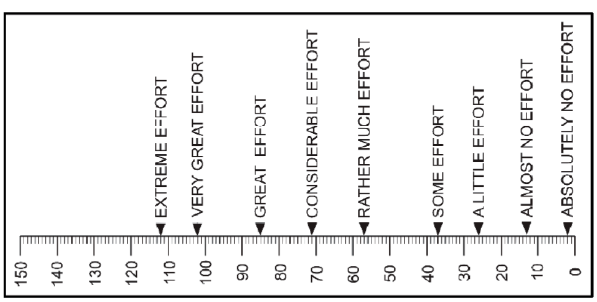

Figure 6-1. Rating Scale Mental Effort (RSME).

Actigraphy

Actigraphy has been proposed as an objective measurement technique for assessing physical activity as a proxy for physical workload, since these wrist-worn devices are able to measure physical movement (Lee & Suen, 2017; McDevitt et al., 2021). Although it has been found that actigraphs can underestimate physical workload in situations with high(er) physical activity such as running, they can determine the level of low and medium physical activity (e.g. as a result of walking) well (Barreira et al., 2009; Herman Hansen et al., 2014), for which it should also be possible to measure physical acticity in cabin crew.

Fatigue related factors

A.1.1 Flight schedules and work-related inputs

The goal of this project is to investigate the effectiveness of the current European Flight Time Limitations, which translate into specific flight schedules for aircrew. Since there can be a difference between the planned and actually flown rosters, the data collection should focus on the rosters achieved by the aircrew that is studied. The most reliable and complete way to gather this information, is to collect the schedule details of the airlines involved, per participating aircrew member. Since it was shown in the previous project that this information can be difficult to collect, another possibility is to gather the information from the participants themselves, by means of digital logbooks. This procedure was successfully used during the previous FTL contract. In addition to information on flight schedules (e.g. departure time, destination, arrival time etc), it is important to gather other relevant work-related data regarding the specific FDPs of interest of this study, such as the number of sectors, the number of time zones crossed (≥4), and the number of crew members aboard, and hassle factors.

A.1.2 Chronotype

Some people would classify themselves as typical night owls or morning larks, reflected in being more active in the (late) evenings or (early) mornings. These differences are in part based on genetic variations that can shorten (resulting in feeling sleepy in the evening, typically a morning lark) or lengthen the circadian period (resulting in feeling sleepy in the morning, typically a night owl) (Patke et al., 2017). These inter-individual differences in circadian period result in different chronotypes, calledearly, intermediate, and late types. The estimated distribution between these types is 30%, 30%, and 40% respectively (Roenneberg et al., 2019). It appears that late chronotypes are more prone to fatigue due to irregular duty schedules than early or intermediate chronotypes (Jung & Lee, 2015; Storemark et al., 2013; Saksvik et al., 2011). Although it is quite difficult to measure chronotype objectively (e.g. through genetics or hormones), several validated questionnaires exist (Horne & Ostberg, 1976; Shahid et al., 2011). Most of theseassume a regular sleep/wake rhythm and do not take irregular work schedules into account. About 10 years ago however, a specific chronotype questionnaire for non-day workers (MCTQshift) has been developed and evaluated (Juda et al., 2013). Another valid possibility to determine one's chronotype is to ask people which chronotype they consider themselves to be, using the self-assessment item from the rMEQ scale (Loureiro & Garcia-Marques, 2015).

Annex B Baseline Questionnaire

Informed consent

I. Informed consent

If you decide to participate, the ethics regulations prescribe that this must be confirmed by clicking "yes" in this Informed Consent Form. With this you indicate that you have understood the information and consent to participation in the study.

Personal Details

II. What is your full name?
III. What is your date of birth?
IV. What is your (professional) email address?

Demographics

V. What is your gender?
a) Female
b) Male
c) Other
d) Prefer not to say

VI. What is your height (in cm)?

VII. What is your weight (in kg)?

VIII. What is your dominant hand?
a) Left-handed
b) Right-handed
c) Ambidextrous

Occupation

IX. What is your occupation?
a) Cockpit crew
b) Cabin crew

A. What is your function?
a) Captain
b) First officer
c) Second officer or Co-Co
d) Other, namely

B. What is your function?
a) Chief purser
b) Purser
c) Flight attendant
d) Other, namely

X. How many years of experience as a commercial air transport (CAT) pilot do you have in total?
XI. How many years of experience as a commercial air transport (CAT) cabin crew member do you have in total?

XII. Are you working fulltime or parttime?
a) fulltime (go to question XIV)
b) parttime

XIII. What is your contract percentage?
a) [open]

XIV. Do you have an additional function (e.g. instructor, manager)?
a) yes
b) no (go to question XVI)

XV. What percentage of your contract do you work in this additional function?
a) [open]

XVI. What is the location of your home base?
a) Northern Europe (Denmark, Finland, Iceland, Norway, Sweden)
b) Eastern Europe (Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, Slovenia)
c) Western Europe (Austria, Belgium, France, Germany, Ireland, Liechtenstein, Luxembourg, Netherlands, Switzerland)
d) Southern Europe (Croatia, Cyprus, Greece, Italy, Malta, Portugal, Slovenia, Spain)
e) Other

XVII. What is the name of the airline you work for?

XVIII. How do you usually commute between home and your homebase?
a) Public transport
b) taxi
c) Car or motorbike
d) Walking or biking
e) Other, namely

XIX. What is your average commute time (in minutes) from home to your homebase?
………. Minutes

Sleep habits

XX. One hears about “morning-types” and “evening-types.” Which one of these types do you consider yourself to be?
a) Extreme morning-type
b) More morning than evening-type
c) Neither morning nor evening-type
d) More evening than morning-type
e) Extreme evening-type

XXI. What is your habitual sleep length? This is the typical time you spend asleep after being moderately tired.
……………………hours …………….. minutes

XXII. How would you rate your sleep quality in general?

a) Very good
b) Fairly good
c) Neither good nor poor
d) Fairly poor
e) Very poor

Sleep duration

XXIII. On work-free days, at what time do you usually fall asleep?

XXIV. On work-free days, at what time do you usually wake up (not using an alarm clock)?

XXV. How long do you usually sleep on workdays? If you have varying working times, please give a rough estimate.
……………………hours …………….. minutes

Health

XXVI. In general, how would you rate your health?
a) Excellent
b) Very good
c) Good
d) Fair
e) Poor

XXVII. Work ability means having the health required for completing your job assuming that the tasks and the environment in which you operate are reasonable. If, at its best, it has a value of 10 points and at its worst a value of 0, how many points would you give your current work ability?
………. Points

Lifestyle

XXVIII. On an average working day, how many caffeinated drinks do you consume (e.g. coffee, tea, energy drinks)?
a) 0
b) 1-2
c) 3-4
d) 5-6
e) > 7

Previous Workload

XXIX. During the **last month**, on average, how well rested did you feel?
a) Fully exhausted
b) Fairly exhausted
c) Slightly exhausted
d) Not exhausted nor rested
e) Slightly rested
f) Fairly rested
g) Fully rested

XXX. During the **last month**, how would you rate your average workload in comparison to your "regular" level?
a) Much lower than regular

b) Somewhat lower than regular
c) Regular
d) Somewhat higher than regular
e) Much higher than regular

Kick-off Meeting

To participate, you will need to attend a kick-off meeting with the other participants, during which you will be briefed on the study procedures and will receive the materials needed for the study. These include a dedicated phone and an ActiWatch. To ensure the date is suitable for as many participants as possible, please provide your availability on the dates below.

XXXI. On what dates are you available for the kick-off meeting?
[List of possibilities]

Schedule

XXXII. On the bulletin that informed you about this study, there is an indicated time period during which the study will take place for your airline. Do you already know your work schedule for this period?
a) Yes > to next question
b) No > end of questionnaire

XXXIII. In this study we are interested in the effect on fatigue of following flights and/or sectors [specific flights/destinations]. Are you scheduled to fly on one or multiple of these flights/Flight Duty Periods in the study period?
a) Yes
b) No

XXXIV. If possible, please upload (a screenshot of) your work schedule for the study period. Please note that this is not mandatory.
[File upload]

Thank you for filling in this questionnaire. We will be in touch as soon as possible on the email address provided by you to give you further information about your participation.

European Union Aviation Safety Agency

European Union Aviation Safety Agency

Konrad-Adenauer-Ufer 3

50668 Cologne

Germany

Mail EASA.research@easa.europa.eu

Web www.easa.europa.eu

An Agency of the European Union