Effectiveness of Flight Time Limitation (FTL)

D2.1 *Identification of Potential Fatigue Hotspots*

Based on survey and roster data

MOVE/C2/2016-360
Classification: Restricted

Table of Contents

Chapter 1: Introduction to the research study

Main objective and scope of the research study

The European Aviation Safety Agency (EASA) was mandated to perform a continuous review of the effectiveness of the rules concerning flight and duty time limitations and rest requirements contained in Annexes II and III of Commission Regulation (EU) No. 965/2012[1].

The review commenced in 2017 with the commission of a research study.

The research study was broken down into smaller phases; each focused on specific flight duty periods (FDPs). The first and current research phase studied the following two FDPs:

- FDP1: Duties of more than 10 hours at the less favourable time of day.
  This focuses on operations that encroach (fully or partially) any portion of the period between 02:00h and 04:59h; and
- FDP2: Disruptive schedules.
  This focuses on consecutive early duty starts, late duty finishes, night duties, and combinations thereof.

Scope of the current deliverable

This Deliverable D2.1 (Identification of Potential Fatigue Hotspots) reports the results of the work performed to identify potential fatigue hotspots in the target population, based on an online survey across Europe and the analysis of historical pilot and cabin crew roster data using two bio-mathematical models.

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

Chapter 2: Approach in identifying fatigue hotspots

This section provides a rationale specifying the criteria for the identification of the fatigue hotspots. We used an online survey and bio-mathematical modelling to identify fatigue hotspots. The survey of pilots and cabin crew mainly served to help interpret the results from the bio-mathematical models. The pilot and cabin crew roster data were gathered from the participating airlines. These rosters were analysed using two bio-mathematical models.

Survey

We developed and used an online survey to identify potential work patterns that were associated with fatigue. That same survey was used to collect crew insights about perceived fatigue hotspots. The respondents selected, from a pre-defined list of 'fatigue items', the items that they deemed to be most relevant for causing the fatigue hotspot. The respondents could also describe in their own words (i.e., answering open questions) how the rosters affect their fatigue, when they feel most fatigued during the duty, and which conditions worsen their fatigue.

The total number of crew respondents was 15,680 (28.4% female); i.e., approximately 10.6% of the entire crew population base in Europe[2]. Of these respondents 58.2% were pilots (4.5% female) and 41.8% cabin crew members (61.5% female). The mean age of all crew respondents was 41 years and 8 months old (range 17-75). The mean age for pilots was 42 years and 4 months and for cabin crew 40 years and 10 months. Of the crew respondents 27.5% indicated to work for a point-to-point operator; 61% worked for a network operator; 3.3% for a cargo operator; and 8.2% for another type of airline.

Roster

Data on planned and worked rosters of pilots and cabin crew members, spanning approximately one year, were collected from the airlines participating in the data collection. In total rosters of six airlines (with a total of 264,746 FDPs) were used for the data analysis. The months of July to October showed the highest numbers of FDPs in the roster dataset. Looking at the geographical distribution and types of operations included in the dataset, a lack of rosters from the northern region of Europe was observed[3].

Airline rosters were analysed using two bio-mathematical models in order to predict the potential level of fatigue; Boeing Alertness Model (BAM[4]) and Sleep, Activity, Fatigue, and Task Effectiveness, Fatigue Avoidance Scheduling Tool (SAFTE-FAST[5]). BAM predicts alertness on common alertness scale (CAS) from 0 (least alert state) to 10,000 (most alert state). CAS is linearly mapped against the Karolinska Sleepiness Scale[6] (KSS) where a KSS value of 9 (very sleepy, great effort to keep awake, fighting sleep) maps to 0 CAS points and KSS of 1 (extremely alert) maps to 10,000 CAS points. For each pilot and cabin crew member BAM was configured to assign a single

[2] As estimated in D2.2 (Definition of the Data Collection Process).
[3] Check D2.2 (Definition of the Data Collection Process) for an estimation of the size and geographical distribution of the entire crew population base.
[4] Jeppesen – as member of the project team – provided access to BAM.
[5] The project team was provided free access to the SAFTE model.
[6] KSS is a 9-point scale: 1. Extremely alert, 2. Very Alert, 3. Alert, 4. Rather alert, 5. Neither alert nor sleepy, 6. Some signs of sleepiness, 7. Sleepy, but no difficulty remaining awake, 8. Sleepy, some effort to keep alert, 9. Very sleepy, great effort to keep awake, fighting sleep.

alertness prediction at top of descent (TOD). TOD was defined at half an hour before wheels on ground.

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

We performed the analysis using the dependent variables CAS and Effect estimated for TOD during the final sector of the FDP. The data analysis plan consisted of the following steps.

Step 1: Check for high predicted fatigue scores

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

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

Step 2: Find clusters of variables

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

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

- The online survey findings;
- The parameters in the bio-mathematical models that were used for the analyses of roster data;
- Scientific literature review[8]; and
- Ideas and suggestions from scientific committee and consortium members.

Step 3: Compare planned and worked rosters

The planned roster data was compared with the worked roster data to study the stability of the roster planning and the impact of roster changes on level of predicted fatigue. In order to be able to compare the planned and worked rosters, we paired (or matched) the datasets using flight numbers, departure locations, arrival locations, and departure dates.

---

[7] SP is a 7-point scale: 1. Fully alert, wide awake, 2. Very lively, but not at a peak, 3. Okay, somewhat fresh, 4. A little tired, less than fresh, 5. Moderately tired, let down, 6. Extremely tired, very difficult to concentrate, 7. Completely exhausted, unable to function effectively.

[8] As presented in D1 (Definition of the Baseline).

Chapter 3: Mapping the identified fatigue hotspots

FDP1 (Night duties of more than 10 hours)

Check for high predicted fatigue scores

We sought to assess the prevalence of high predicted fatigue scores during duties of more than 10 hours encroaching on the night. Our results show that – according to the bio-mathematical models used to examine the roster data – high fatigue scores did occur in flight duties longer than 10 hours that encroached partially or fully on the period between 02:00h and 04:59h. Moreover, the proportion of high fatigue was greater in these FDPs than in the baseline set containing all the collected FDPs. This was confirmed by the survey results, as 14% of respondents indicated 'a long working day' and 11% of respondents indicated 'flying during hours when I would normally sleep' as fatigue factors, and both were linked to night duties of more than 10 hours.

Find clusters of variables

We sought clusters of FDP-related characteristics that might impact the level of predicted fatigue during night FDPs of more than 10 hours. We worked our way towards multiple logistic regression models for the CAS and Effect measures, estimated for TOD during the final sector of the FDP.

The resulting multiple regression models (based on either the CAS or Effect dependent variable) differed to some extent, which can only be the result of differences between the bio-mathematical models used, as the same datasets were used in the analysis. Although both models are 'two-process models', they are not alike in how they represent and implement the two processes; i.e., the mathematical representation of each process is different, the relative weighting of the two processes is different, and the manner in which the two models estimate a pattern of sleep associated with a sequence of duties is different. The two models trace their roots back to different research data (either primarily validated against alertness ratings or cognitive performance). This difference may account for the differences in weightings of fatigue factors.

The resulting FDP1 multiple regression models (also referred to as clusters of variables) included the following predictors:

- Sleep prediction in 24h prior TOD of the final sector;
- Start and end time of the FDP;
- Time in (alternative) window of circadian low (WOCL: 02:00h - 05:59h); and
- Number of time zones crossed east- and westwards.

The predicted duration of sleep in the 24 hours preceding the FDP was included in the multiple regression model because sleep is the primary recovery mechanism for fatigue. The bio-mathematical models differ in the way they estimate sleep; e.g., CAS includes no consideration of the effect of sleep on the odds of high fatigue. There is, however, an indirect link, as 23% of survey respondents indicated 'insufficient time between duties' as one of the contributing factors to fatigue. Insufficient time between duties may be interpreted as not enough time to get a good sleep.

For night FDPs of more than 10 hours, the earlier start and later end time of the duty period were included in the multiple regression models. This was not surprising, as the body's circadian rhythm has a major effect on fatigue levels. This is confirmed by the significance of the WOCL variable. Here, we see a clear link with the survey results, as 39% of respondents indicated that they were most fatigued 'in the WOCL', and 22%

replied 'at the end of the duty'. Nine percent of the respondents indicated 'flying or being awake during hours when I would normally sleep' as a factor contributing to fatigue. Note that in the survey fatigue was not reported specifically at TOD, as opposed to in the roster data analyses.

For night FDPs of more than 10 hours the number of time zone crossings was included in the models because crossing more time zones results in longer FPDs that potentially encroach (part of) the WOCL. There was only a limited reference to time zone crossings in the survey responses; i.e., in- and outbound flights crossing more than six time zones was indicated in around 3% of the cases.

Relevant rules in Subpart FTL

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

- Start of the FDP at reference time (ORO.FTL.205 FDP);
- Sleep opportunity in 24h and 48h (ORO.FTL.205 FDP and ORO.FTL.235 Rest Periods);
- Flight duration (ORO.FTL.205 FDP and ORO.FTL.210 Flight Times and Duty Periods);
- Number of time zones crossed (ORO.FTL.235 Rest Period); and
- Duty time in WOCL (ORO.FTL.205 FDP).

FDP2 (Disruptive schedules)

Check for high predicted fatigue scores

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

For the *non-consecutive* disruptive schedules we found a relatively high prevalence of fatigue for the late finishes and nights. This was underpinned by the survey results, as 7% of respondents indicated 'late finishes' and 11% indicated 'flying during hours when I would normally sleep' as causes of fatigue. This was not the case for early starts, for which the prevalence of fatigue was low. In somewhat of a contradiction, a relatively high percentage of survey respondents (12%) indicated 'starting early' as a relevant fatigue item. However, it should be noted here that the survey questions used 'fatigue' as a broad term including physical fatigue, mental fatigue, and sleepiness and that the survey questions were not specifically focused on TOD. Nonetheless, the proportion of high fatigue was greater for the non-consecutive disruptive duties than for the baseline set, except for early starts, and the proportion for late finishes was just below the 1.0 relative ratio for Effect.

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

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

Find clusters of variables

We sought clusters of FDP-related characteristics that might impact the level of predicted fatigue during *consecutive* disruptive flight duties. Similar to the approach used for FDP1 (Night duties of more than 10 hours) we worked our way towards multiple logistic regression models for the CAS and Effect measure. The conclusions drawn below are based on the independent variables included in the resulting multiple regression models.

The resulting multiple regression models per consecutive disruptive duty differed to some extent which results from the differences between the bio-mathematical models.

Consecutive early starts

Multiple regression models for consecutive early starts could not be computed because the prevalence of high predicted fatigue was very low; i.e., a prevalence of 1 (CAS) and 0 (Effect) in 6,895 observations. The survey results indicated 'consecutive early starts' to describe how the preceding roster affected their fatigue in 4% of the cases.

Consecutive late finishes

Only for the CAS measure could a multiple logistic regression model be developed as there were no valid cases for modelling of the Effect measure. The resulting FDP2 model for consecutive late finishes included the following predictors:

- Late finishes;
- Sleep opportunity in 48h in darkness prior TOD of the final sector;
- FDP duration; and
- Time awake prior TOD.

The variable very late finish shows a relatively strong increase in odds of high fatigue. The survey results indicated 'finishing late' as a relevant fatigue item in about 7% of the cases.

The fact that duration of night-time sleep opportunity in the 48 hours preceding the FDP was included in the multiple regression model was expected, because sleep is the primary recovery mechanism for fatigue. There is an indirect link to the survey result that 33% indicated 'insufficient time between duties' as one of the contributing factors to fatigue.

With regard to FDP duration, too, the survey results provide a link. Some 8% of respondents indicated 'long working days' as a contributor to fatigue. This was confirmed in the model by the significance of an extended time awake prior to TOD. In addition, 32% of survey respondents indicated 'at the end of the duty' as answer to the question 'when do you feel most fatigued'.

Consecutive night flights

The resulting FDP2 multiple regression models for consecutive nights included to following predictors:

- Time in (alternative) WOCL;
- Start and end time of the FDP; and
- FDP duration.

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

WOCL was identified as a predictor in the model for the Effect measure. The importance of this variable appeared in the survey results as well. In response to the question 'when do you feel most fatigued', 38% of respondents indicated 'in the WOCL' as one of their answers. Seven percent of the respondents indicated 'flying or being awake during hours when I would normally sleep' as a factor contributing to fatigue; and 8% indicated 'unfavourable time for resting' as a factor.

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

Survey respondents (8%) indicated 'a long working day' as a contributor to high fatigue in disruptive flight schedules. This variable (i.e., FDP/flight duration) was also included in the multiple regression models and seems to be related to the end and start hour of the FDP.

<u>Mix of early starts, late finishes and night flights</u>
The multiple regression models for a mix of disruptive schedules included the following predictors:

- Early starts and late finishes;
- Transitions from late finish to night;
- Time in WOCL;
- Start and end of FDP; and
- Time of day.

The variables early start and (very) late finish were included in the CAS model; these variables overlap with the grouping variable for mix of disruptive schedules. Given the grouping variable, it also makes sense that several WOCL variables were included as predictors. The same goes for the transitions from late finish to night duties.

Sleep was not identified as a predictor in either model. Later start and end time of the duty period and later time of day increase the odds for high fatigue.

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

- Start of the FDP at reference time (ORO.FTL.205 FDP);
- Sleep opportunity in 24h and 48h (ORO.FTL.205 FDP and ORO.FTL.235 Rest Periods);
- Flight duration (ORO.FTL.205 FDP and ORO.FTL.210 Flight Times and Duty Periods);
- Number of time zones crossed (ORO.FTL.235 Rest Period);
- Duty time in WOCL (ORO.FTL.205 FDP); and
- Rest period provided before undertaking an FDP (ORO.FTL.235 Rest Period).

Planned vs worked rosters

The goal of this step was to check the stability of the roster planning and the possible impact of roster changes on level of predicted fatigue. Relevant to mention is that 15% of the survey respondents indicated 'delays' as a condition that may worsen fatigue. The item 'changes in the schedule' was indicated in 10% of the time.

The roster analysis did not result in any evidence for the level of fatigue being impacted by delayed arrivals. As the datasets were completely de-identified before delivery, we could not determine on an individual crew level whether or not a (late) change in flight schedule had a worsening effect on predicted sleep and level of fatigue.

List of abbreviations

Abbreviations	***Description***
BAM	Boeing Alertness Model
CAS	Common Alertness Scale
D	Deliverable
EASA	European Aviation Safety Agency
EC	European Commission
EU	European Union
FAST	Fatigue Avoidance Scheduling Tool
FDP	Flight Duty Period
FTL	Flight Time Limitation
KSS	Karolinska Sleepiness Scale
ORO	Organisation Requirements (in the air Operations Regulation)
SAFTE	Sleep, Activity, Fatigue, and Task Effectiveness
SP	Samn-Perelli
TOD	Top Of Descent
WOCL	Window Of Circadian Low

European Commission