Effectiveness of Flight Time Limitation (FTL)

***D2.3** Performance of the Data Collection and Data Analysis*

Based on field data

MOVE/C2/2016-360
Classification: Restricted

Table of Contents

Chapter 1: Introduction to the research study 3
Main objective and scope of the research study 3
Scope of the current deliverable 3
Chapter 2: Approach in identifying fatigue hotspots 4
Step 1: Check for high fatigue scores 4
Step 2: Compare fatigue scores between FDP categories 4
Step 3: Find clusters of variables 5
Chapter 3: Mapping the identified fatigue hotspots 6
Crew data representativeness and sample size 6
The main analyses results 6
Fatigue at TOD during the FDPs of interest 6
Predicting high fatigue at TOD during the FDPs of interest 10
Additional analysis with special reference to FDP duration 11
Additional results of alternative FDP categories 11
Summary of the results 14
Discussion and conclusions 15
Understanding the main analyses results 15
Implications for fatigue mitigation 16
Need to revise regulations? 17
Chapter 4: Critique of the whole data collection activity 18
Critique assessment 18
Resulting critique of the data collection 18
Recruitment and training of crew members 18
Data collection tools 18
References 20
List of abbreviations 21

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 D2.3 (Performance of the Data Collection and Data Analysis) reports on the data collection and analysis. The aim of the work was the identification of potential schedules and FDPs likely to be associated with high fatigue within the target population.

[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 for the identification of the fatigue hotspots and explains the data analysis. Fatigue hotspots are defined as schedules that are associated with high on-duty fatigue.

Crew members from participating airlines[2] collected data for approximately two weeks (per participant) between July 2017 and February 2018.

In line with the approach defined in D2.2 (Definition of the Data Collection Process), the primary data analyses were performed using the KSS at top of descent (TOD) during the final leg of the FDP. The data analysis plan consisted of the following steps.

Step 1: Check for high fatigue scores

The goal of this step was to identify whether or not high fatigue scores occurred in FDP1 (Night duties of more than 10 hours) and FDP2 (Disruptive schedules).

A high level of fatigue was defined by scores on the Karolinska Sleepiness Scale (KSS)[3] ≥ 7 and Samn-Perelli (SP)[4] scale ≥ 6. Total sleep in 24 hours prior to TOD (Sleep24h) and Napping during the FDP (FDPsleep) for the high (KSS ≥ 7) and low levels of fatigue (KSS < 7) was described. In addition, the percentages of high and low fatigue scores (KSS) for each hour of the day were presented. The reason for the selection of KSS = 7 as cut-off is that numerous studies have shown that, at and above this score, performance levels start to decrease (e.g. Åkerstedt et al. 2014).

Step 2: Compare fatigue scores between FDP categories

Differences were calculated between the KSS scores in FDP1 (night duties of more than 10 hours) and those in control FDP categories, and between FDP2 (disruptive schedules) and control FDP categories.

The main analysis approach for FDP1 (Night duties of more than 10 hours) was to compare short FDPs (≤ 10h) with long (> 10h) ones, with respect to level of fatigue at TOD, and adjustment for factors that may influence the outcome. The same analysis was repeated using, instead of a cut-off of 10h, cut-offs of 8 hours, 9 hours, 11 hours, and 12 hours.

The approach for FDP2 (Disruptive schedules) was to compare all disruptive FDPs with all non-disruptive (essentially daytime) FDPs. Additional comparisons were performed for the different types of FDP2 (early starts, late finishes, and nights) and between one disruptive FDP and two successive disruptive FDPs.

---

[2] D1 Addendum provides an overview of the candidate airlines for the data collection.

[3] KSS is a 9-point scale: 1. Extremely alert, 2. Very Alert, 3. Alert, 4. Rather alert, 5. Neither alert nor sleepy, 6. Some sighs 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.

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

Step 3: Find clusters of variables

The goal of this step was to develop multiple logistic regression models that could be used to determine clusters of FDP-related characteristics (or independent variables) under which high levels of fatigue occur, also referred to as fatigue hotspots.

Variables that may contribute to fatigue were defined based upon the following sources:

- The online survey findings[5];
- The parameters in the bio-mathematical models that were used for the analyses of roster data[6];
- Scientific literature review[7]; and
- Ideas and suggestions from scientific committee and consortium members.

[5] As presented in D2.1 (Identification of Potential Fatigue Hotspots).
[6] As presented in D2.1 (Identification of Potential Fatigue Hotspots).
[7] As presented in D1 (Definition of the Baseline).

Chapter 3: Mapping the identified fatigue hotspots

Crew data representativeness and sample size

The 24 airlines participating in the data collection were used as a reference set for the EU aviation sector as a whole. Our examination of these participants suggested that the eastern region might be overrepresented in the dataset, with six airlines participating. However, these six airlines are small relative to the others. Thus, based on the geographical distribution and type of operations included, we consider it representative. It is thus appropriate to use the set as a proxy for the EU aviation sector. This conclusion is confirmed by our estimates of the size and geographical distribution of the entire EU aircrew population as described in D2.2 (Definition of the Data Collection Process).

Data was collected by 381 crew members[8] and for 2877 FDPs. The participating crew population consisted of 68% pilots and 32% cabin crew, whereas in the entire EU crew population approximately 59% are cabin crew[9]. One reason for the relatively higher proportion of pilot participation in the data collection is that three participating airlines were cargo operators (i.e., in line with the need to include operators that operate disruptive/night FDPs on a regular base). The cargo operators only had a small number of cabin crew (2 cabin crew members), presumably because there are just a small number of cabin crew employed by the operators.

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

The main analyses results

Fatigue at TOD during the FDPs of interest

Figure 1 presents point estimates of the occurrence probability of high fatigue at TOD (KSS ≥ 7) during the FDPs of interest and in the entire dataset (i.e., all FDPs collected – referred to as the baseline). Compared to the baseline, the point estimates clearly increased for long and short night FDPs (duration > 10h and ≤ 10h), for night FDPs (including all night FDPs), and for late finishes FDPs. A marginal increase was found for the mixed FDPs. The mixes represented the following combinations: an early start FDP preceded by a late finish FDP, an early start FDP preceded by a night FDP, a late finish FDP preceded by an early FDP, a late finish FDP preceded by a night FDP, a night FDP preceded by an early start FDP, and a night FDP preceded by a late finish FDP. No increase was found for early start FDPs. Keep in mind that these results are descriptive only; i.e., they do not represent a statistical comparison between the FDPs of interest and the baseline. Also note that the baseline represents the mean level across all FDPs (not, e.g., daytime FDPs only).

[8] That is approximately 0.3% of the entire crew population base in Europe as estimated in D2.2 (Definition of the Data Collection Process).
[9] As described in D2.2 (Definition of the Data Collection Process).

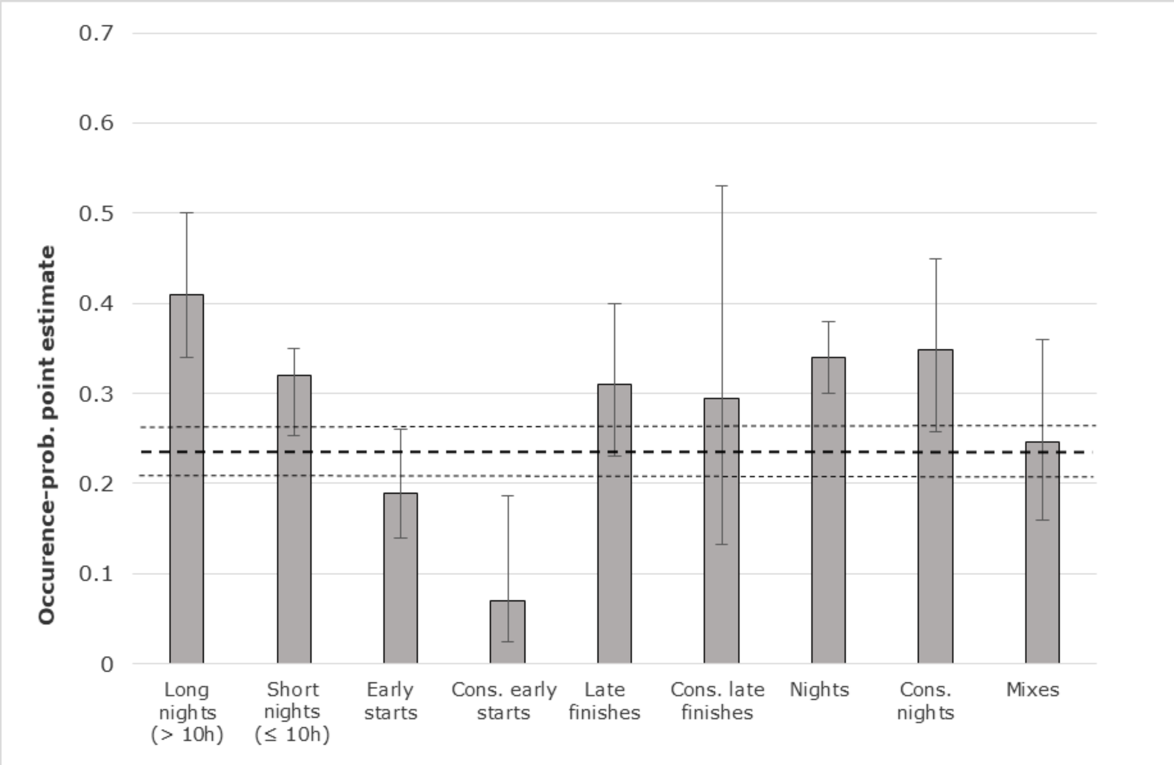

Figure 1 Point estimates for the occurrence probability of high fatigue at TOD during the FDPs of interest and the baseline condition (all FDPs collected and denoted by the thick dashed horizontal line). The thin dashed lines denote the 95% confidence interval (CI). The vertical lines indicate 95% CIs of the FDPs of interest. The number of observations by FDP type are as follows: long nights (> 10h) 146, short nights (≤ 10h) 348, early starts 163, consecutive early starts 43, late finishes 123, consecutive late finishes 17, nights 494, consecutive nights 92, and mixed combinations of disruptive schedules 69

Figure 2 shows odds ratios (ORs) for high fatigue (KSS ≥ 7) at TOD during the FDPs of interest. These analyses were based on between-subjects data extracted from the entire dataset (baseline) and the reference used was all daytime FDPs (all FDPs with start time ≥ 07:00h and end time < 23:00h).

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

Keep in mind that the data presented in Figure 2 do not allow fair comparisons to be made between the FDPs of interest, but only between the FDP of interest and the reference condition. Also note that the reference condition for long-night FDPs is short-night FDPs, unlike the other FDPs of interest, which use daytime FDPs as their reference.

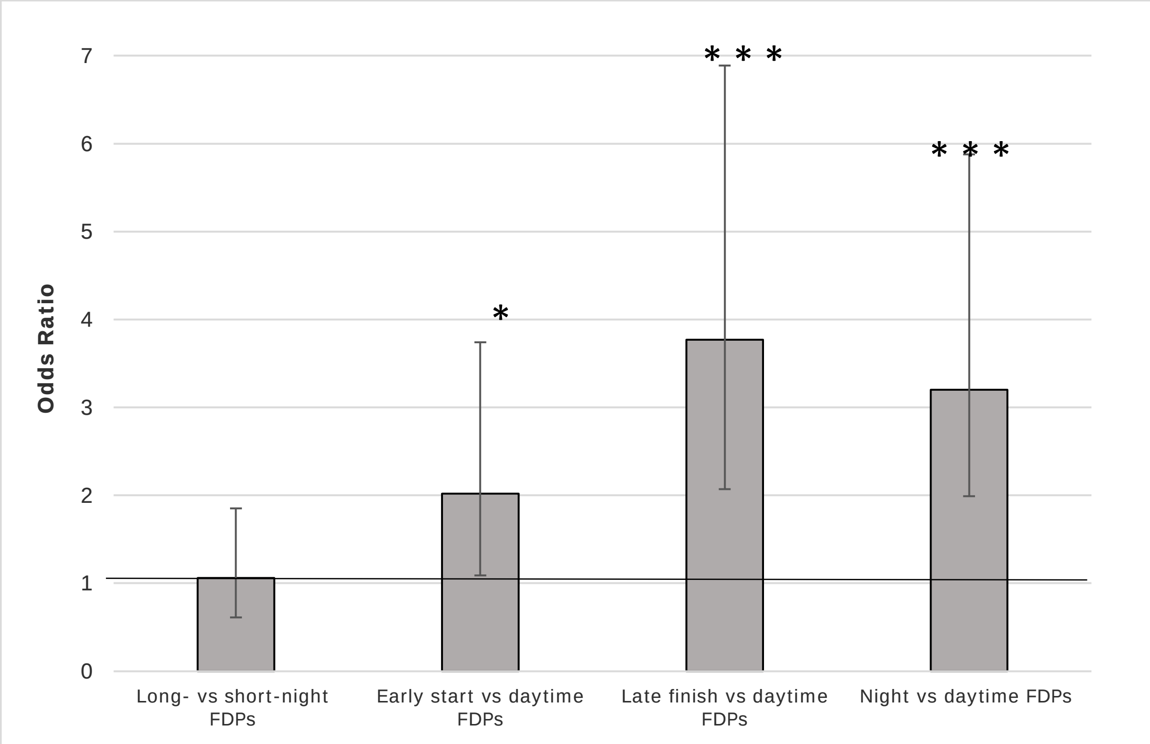

Figure 2 ORs for reporting high fatigue at TOD (KSS ≥ 7) during the FDPs of interest compared to their reference conditions. The horizontal line denotes the reference FDP category. Note that the reference condition for long-night FDPs is short-night FDPs, unlike the other FDPs of interest, which use daytime FDPs as their reference. The vertical lines indicate the 95% CIs. A value greater than 1 indicates an increased OR. * = $p < 0.05$; *** = $p < .001$

Figure 3 shows the results of a supplementary analysis comparing long- and short-night FDPs in more detail using a between-subjects dataset extracted from the baseline. No significant differences in fatigue at TOD were observed for long-night FDPs when the criterion for long duration was varied between > 8 hours and > 12 hours.

Figure 4 shows the mean KSS values at TOD during the first and second FDPs of interest in a row. None of the comparisons within each FDP type (early start, late finish, night) indicated an increase in fatigue (from the 1st to the 2nd FDP in a row).

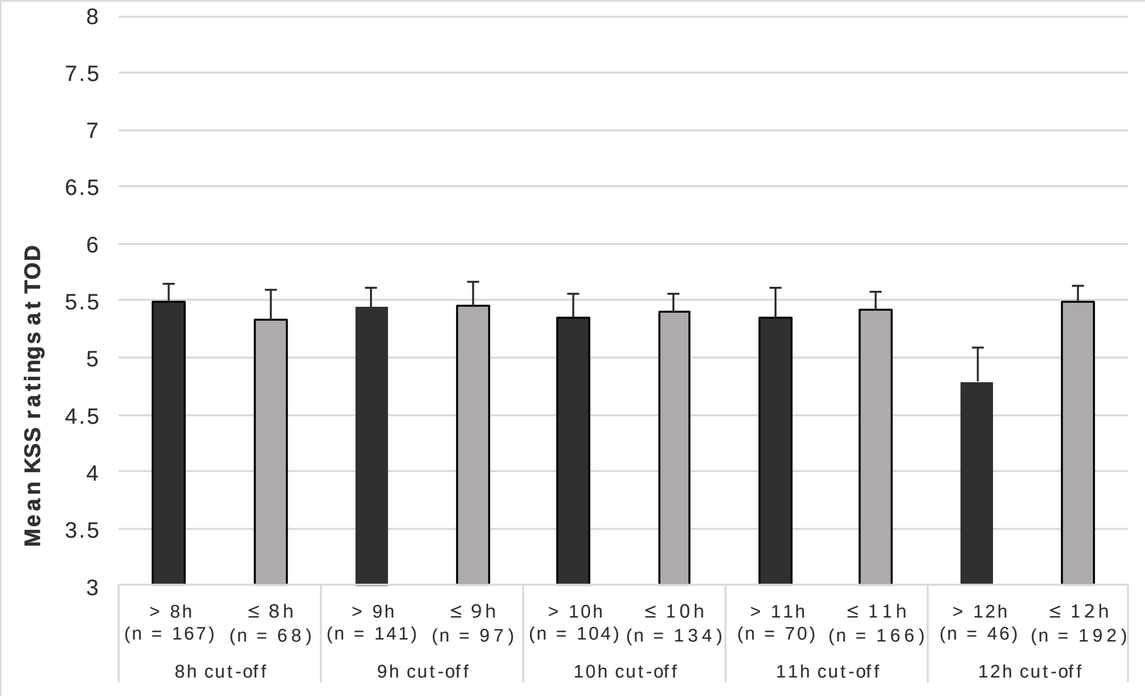

Figure 3 Mean KSS ratings at TOD for the long- and short-night FDPs. The criterion for the duration of a long-night FDP ranged from > 8h to > 12h. The black bars represent long FDPs and the grey bars short FDPs in the between-subjects data. The vertical lines denote the standard errors. Note that the y-axis covers only part of the 9-point KSS

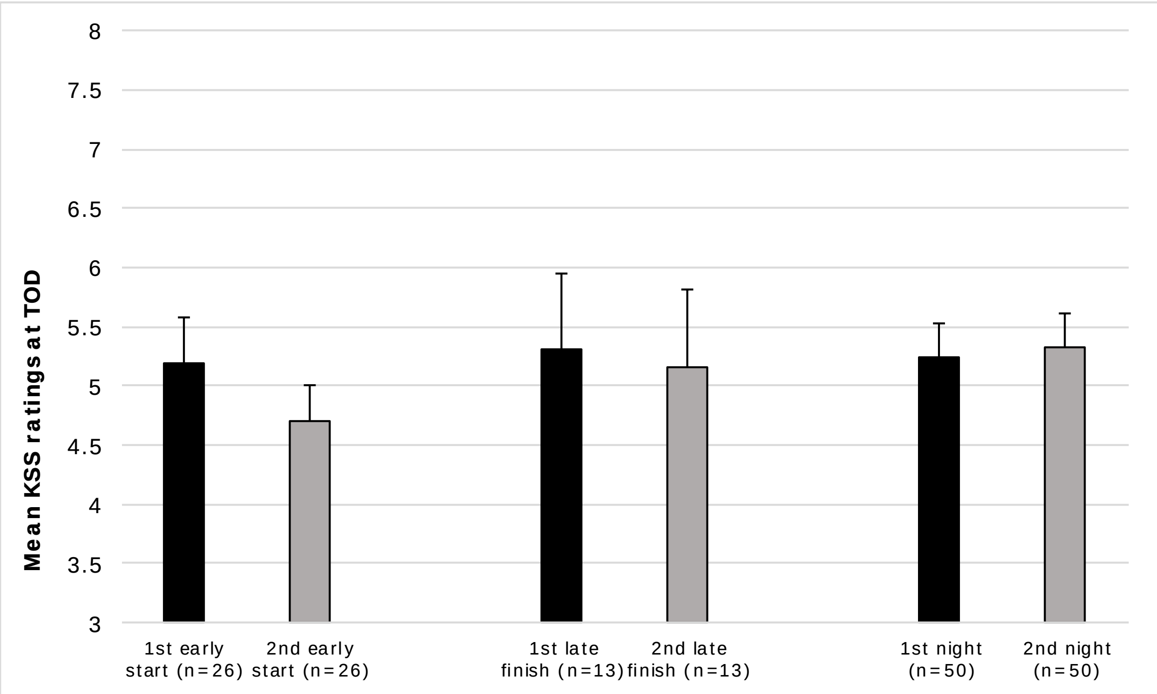

Figure 4 Mean KSS ratings at TOD for the first (black bars) and second consecutive (grey bars) early start (n = 26), late finish (n = 13), and night (n = 50) FDPs in the within-subject data. The vertical lines denote the standard errors. Note that the y-axis covers only part of the 9-point KSS

Predicting high fatigue at TOD during the FDPs of interest

Figure 5 shows the main FDP-related predictors of high fatigue at TOD. The results are based on datasets that include an FDP of interest and its reference FDP (daytime/non-disruptive FDPs). Results of the multiple regression analysis indicate that the increased odds of high fatigue at TOD during early start FDPs are attributable only to the earlier start time itself. When early start FDPs were analysed without their reference condition, none of the FDP-related characteristics (including prior sleep) explained the occurrence of high fatigue at TOD.

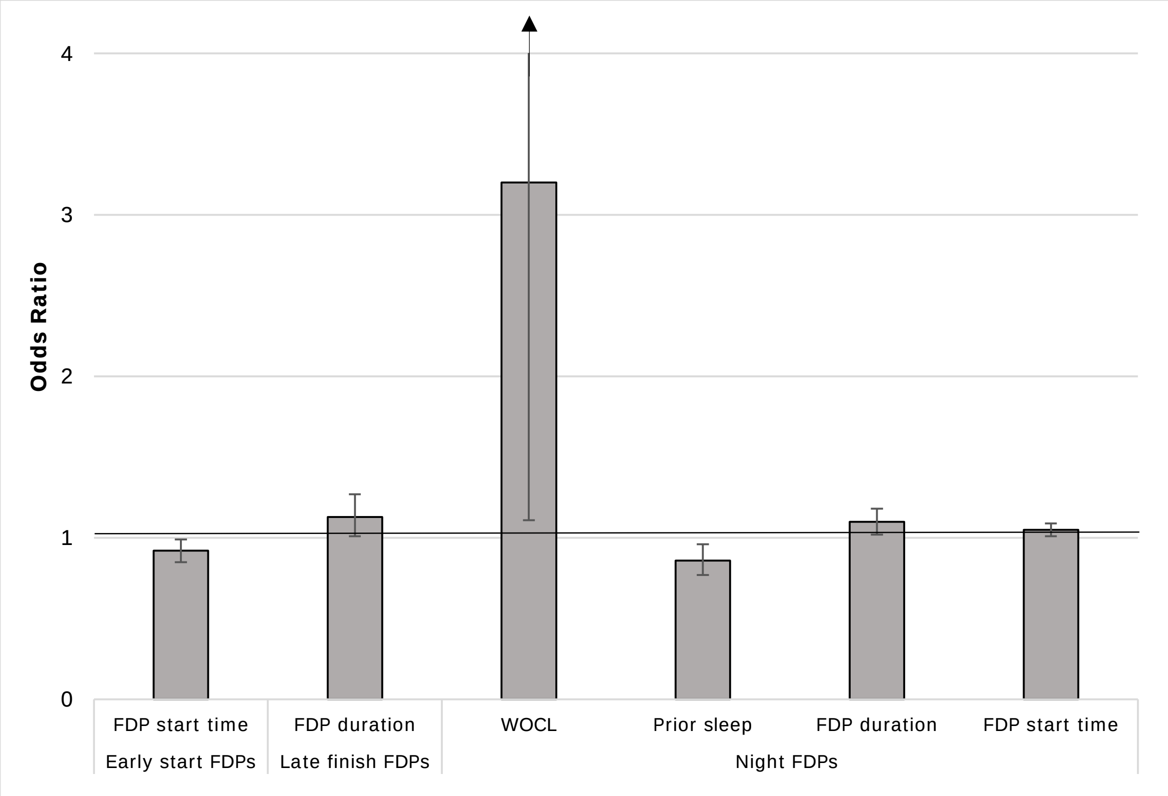

Figure 5 ORs of high fatigue at TOD (KSS ≥ 7) for the FDPs of interest and the reference FDP (daytime). The horizontal line denotes the reference FDP (all FDPs with start time ≥ 07:00h and end time < 23:00h). The vertical lines indicate the 95% CI. A value greater than 1 indicates an increased OR. WOCL = window of circadian low. Note that the upper limit of the 95% CI for the WOCL fall outside the y-axis scale (9.19), which indicated by an arrow

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

For night FDPs, the main predictors of increased odds of high fatigue at TOD were on duty during WOCL (02:00h - 05:59h) and shorter prior sleep in the past 24 hours. Longer sleep acted as a protective factor, as indicated by the OR of < 1 (see Figure 5). After removal of these two factors from the regression model, longer FDP duration and later FDP start time became significant predictors. Besides these FDP-related characteristics, being a cabin crew member (versus being a pilot) was associated with increased odds of high fatigue at TOD for night FDPs (OR 1.76; 95% CI 1.07 - 2.90, $p < 0.05$).

In a supplementary analysis without reference FDPs, only the amount of sleep in the past 24 hours and being a cabin crew member (versus being a pilot) significantly predicted the occurrence of high fatigue at TOD during night FDPs (prior sleep: OR 0.84; 95% CI 0.75 - 0.94, $p < 0.01$; cabin crew member: OR 1.89; 95% CI 1.07 - 3.36, $p < 0.05$). The effect of WOCL could not specifically be analysed because the WOCL was the basis of the definition of night FDPs.

Additional analysis with special reference to FDP duration

As FDP duration showed very limited predictive power regarding high fatigue at TOD for night FDPs, we made an additional attempt to examine this factor in depth. First, we divided night FDPs into two categories using the FDP end time of 06:00h as the cut-off. A regression analysis performed separately for these two night FDP categories did not, however, show FDP duration to be associated with fatigue at TOD (end time < 06:00h: OR = 1.04, CI = 0.86 - 1.27; end time ≥ 06:00h: OR = 0.85, CI = 0.63 - 1.13).

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

Finally, we analysed the relationship between FDP duration and high fatigue at TOD for daytime FDPs (those starting and ending between 07:00h and 23:00h). A simple regression analysis found only FDP duration (OR = 1.24, CI = 1.09 - 1.41, $p = 0.001$) and end time (OR = 1.13, CI = 1.00 - 1.28, $p = 0.045$) to be significant predictors. Entering the two predictors into a multiple logistic regression, we obtained a reduced, but still significant result for FDP duration (OR = 1.21, CI = 1.04 - 1.39, $p = 0.012$). The range in FDP duration was considerable for daytime FDPs (0.75h - 13.8h), which is in contrast to the overall scarcity of short FDPs in the night FDPs. This contrast might contribute to the difference found between night FDPs and daytime FDPs in the association between FDP duration and high fatigue at TOD.

Additional results of alternative FDP categories

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

We made two changes to the classification of the FDPs of interest. First, we created two categories for early start FDPs: early start FDPs (start time between 05:00h and 06:59h) and deep early start FDPs (start time between 02:00h and 04:59h). This division enables us to determine if starting time is a significant factor in early start FDPs. In ORO.FTL, all FDPs starting between 02:00h and 04:59h are considered night FDPs, though the crew is usually able to obtain at least some night sleep just prior to their FDP.

Second, we created two categories for night FDPs: night FDPs (ending between 02:00h and 05:59h) and deep night[10] FDPs (ending 06:00h or later). This division enables us to determine if night FDPs that end within the WOCL differ from those that end after the WOCL (and thus completely cover the WOCL). This new classification is closely linked to the three-process model of alertness, which is a scientifically established model to predict sleep and fatigue (Åkerstedt & Folkard, 1997).

Figure 6 shows point estimates for the occurrence probability of high fatigue at TOD in each of the FDP categories, compared to non-disruptive FDPs (starting 07:00h or later and ending at 22:59h or earlier). Each FDP of interest showed an increased tendency towards high fatigue. This tendency was especially discernible during late finish, night, and deep night FDPs. Among these, the tendency was most pronounced in the deep night category. The results for the early start and deep early start FDP categories were very similar.

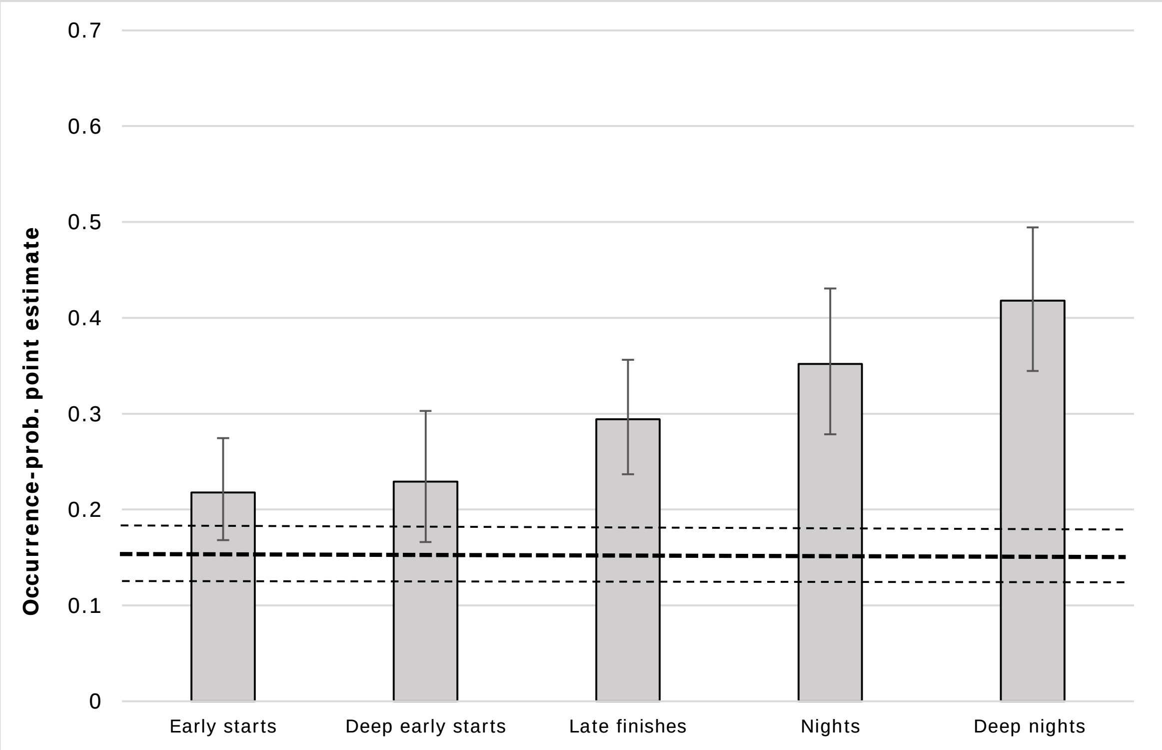

Figure 6 Point estimates for the occurrence probability of high fatigue at TOD in each FDP category of interest. The vertical lines denote a 95% CI. The thick dashed horizontal line denotes the daytime FDP category (all FDPs with start time ≥ 07:00h and end time ≤ 22:59h). The thin dashed lines represent the 95% CI

Why are the late finish and night FDPs particularly fatiguing? One tentative explanation is a reduced sleep-wake ratio. In the FDPs of interest, that ratio fell clearly below the level of the daytime FDPs (mean 0.57; 7.35 hours of sleep followed by 12.85 hours of wakefulness) especially in the two night FDP categories (mean 0.30; 0.22). It also bears mentioning that the deep night FDPs were exceptionally long in duration (mean 10.10 hours), which may also explain the result. A tentative reason

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

for not finding a difference between the early start and deep early start FDPs was the unexpectedly favourable sleep-wake ratio in both categories (mean 0.50).

Table 1 shows logistic regression results for the different FDP categories. In each of these categories, the OR of high fatigue at TOD was increased compared to daytime FDPs. The most pronounced increase was found in the deep night category. In this category, the OR was about two times higher than in the other four categories.

Table 1 Results from logistic regression predicting high levels of fatigue in the alternative FDP categories. The analyses are based on between-subjects data. Daytime FDPs serve as the reference condition

FDP category	***OR***	***CI***	***p***	***N*** *daytime/disruptive*
Night	4.16	2.00;8.65	.000	165/51
Deep night	8.04	3.58;180	.000	154/63
Early	3.28	1.30;8.25	.012	174/39
Deep early start	4.16	1.63;10.22	.000	170/30
Late finish	4.65	2.08;10.40	.000	190/53

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

Summary of the results

These results are exclusively based on the field data collected and are discussed in detail in the section 'Discussion and conclusions'.

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

FDP of interest	***Main results on high fatigue[1] at TOD and its predictors***
Night duties (> 10h)	In the data collected, all night FDPs were associated with high probability of high fatigue at TOD. This probability was similar for short- (≤ 10h) and long-nights (> 10h).
Nights	In the data collected, the probability of high fatigue at TOD during night FDPs was higher compared to during daytime FDPs. Encroachment of the FDP on the WOCL and shorter prior sleep were the significant predictors. When night FDPs were analysed alone, shorter prior sleep explained the occurrence of high fatigue at TOD.
	To cover the continuum from evening to night, late finish plus night (start time before 00:00h) FDPs were combined. Increased probability of high fatigue at TOD during these FDPs compared to during daytime FDPs was predicted by encroachment on the WOCL, shorter prior sleep, later FDP start time, and longer FDP duration. When late finish plus night FDPs were analysed alone, encroachment on the WOCL, earlier FDP end time, and shorter prior sleep explained the occurrence of high fatigue at TOD.
	To cover the continuum from late night to early morning, very early (03:00h - 04:59h) and early (05:00h - 06:59h) starting FDPs were combined[2]. Increased probability of high fatigue at TOD during these FDPs as compared to daytime FDPs was explained by earlier FDP start time and shorter prior sleep. When these FDPs were analysed alone, only shorter prior sleep explained the occurrence of high fatigue at TOD.
	An alternative way of classifying FDPs was suggested. When applying this classification, probability of high fatigue at TOD was found to be similar for deep early (start time 02:00h - 04:59h) and early (start time 05:00h - 06:59h) start FDPs[3]. The highest probability of high fatigue at TOD was found for deep night FDPs that covered the entire night (start time 01:59h or earlier, end time 06:00h or later).
Early starts	In the data collected, the probability of high fatigue at TOD during early start FDPs was higher compared to during daytime FDPs. Earlier FDP start time was the only statistically significant predictor. When early start FDPs were analysed alone, none of the FDP-related characteristics explained the occurrence of high fatigue at TOD.
Late finishes	In the data collected, the probability of high fatigue at TOD was higher during late finish FDPs compared to during daytime FDPs. Longer FDP duration was the only significant predictor. When late finish FDPs were analysed alone, longer FDP duration, earlier FDP start time, and earlier FDP end time explained the occurrence of high fatigue at TOD.

[1] A high level of fatigue was defined by scores on the Karolinska Sleepiness Scale (KSS) scores equal or greater than 7 (= sleepy, but no effort to keep awake).
[2] Note that those FDPs that started between 03:00h and 04:59h are considered as night FDPs in the current FTL whereas those starting between 05:00h and 06:59h are not.
[3] The deep early FDPs are considered night FDPs in the current FTL, whereas early FDPs are not.

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

FDP	***Main results on high fatigue[1] at TOD***
Consecutive early starts	In the data collected, fatigue levels at TOD were similar for the first and second early start FDPs in a row.
Consecutive late finishes	In the data collected, it seemed that fatigue levels at TOD were similar for the first and second late start FDPs in a row.
Consecutive nights	In the data collected, fatigue levels at TOD were similar for the first and second night FDP in a row.
Mix	In the data collected, it seemed that the probability of high fatigue at TOD during mixes of disruptive schedules was higher compared to the corresponding probability in the entire dataset.

[1] A high level of fatigue was defined by scores on the KSS scores equal or greater than 7 (= sleepy, but no effort to keep awake).

Discussion and conclusions

Understanding the main analyses results

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

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

No significant difference in fatigue at TOD was found between night duties of more than 10 hour, compared to shorter night FDPs. Our result does not, however, mean that FDP duration is not an important determinant of fatigue. The main reason for the result probably is that high fatigue during night FDPs is mainly caused by the unfavourable time of the day (circadian factor) and a reduced sleep-wake ratio (homeostatic factor). These two factors likely interacted with the influences of FDP duration. In addition, night FDPs seldom are of short duration (i.e., in the field dataset 1.5% of the night FDPs were found in the ≤ 4h category; in the roster dataset this was 8.6%), which limits the range of variation of this FDP characteristic. It is also important to note that we did not measure the length of time participants were fatigued. This limitation can be assumed to underestimate FDP duration as a factor underlying fatigue in our analyses.

No significant difference in fatigue at TOD was found between the first and second consecutive disruptive FDPs. Unfortunately, our field data did not permit us to study cumulative fatigue over sequences longer than two consecutive FDPs. This might be the result of the current regulatory fatigue management controls and/or company rostering rules. The roster data[11] also showed relatively low sample sizes for the different types of consecutive disruptive schedules. This is especially the case for four or more disruptive schedules in a row. This is likely associated with the required extension of the recovery rest period if a crew member performs four or more disruptive schedules (CS.FTL.1.235 Rest Periods). The same lack of data holds for schedules where an early start FDP is preceded by a duty sequence that compromises sleep (e.g. quick transitions). This limitation restricts our possibilities to explore the fatigue associated with the different types of disruptive schedules.

The strongest predictors of increased probability of high fatigue at TOD, as compared to daytime FDPs, varied by FDP type. For early start FDPs, only earlier start time itself was a significant predictor. For late finish FDPs, the only significant predictor was longer FDP duration. In case of night FDPs, the pertinent predictors were encroachment on the WOCL (02:00h - 05:59h), short prior sleep, and being a cabin crew member.

The difference between pilots and cabin crew is of interest. A likely explanation for the result lies in a difference in the level of workload at TOD between the two crews. At TOD, cabin crew are typically sitting in the cabin crew jump seat after a potentially busy work period in the cabin and in a low workload phase of flight. In contrast, at TOD pilots are in a high workload phase of flight, having just finished preparing for

[11] Presented in D2.1 (Identification of Potential Fatigue Hotspots).

descent, approach and landing and now commencing descent. Unlike role, age was not a significant individual factor, as it showed some predictive power only during daytime FDPs. The other individual factors examined – diurnal type, habitual sleep length, body mass index, and commuting time – were not significant predictors of high fatigue during any FDP type.

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

Implications for fatigue mitigation

Night duties of more than 10 hours

Our results suggest that increased fatigue at TOD during night duties of more than 10 hours may be difficult to effectively control by just adjusting FDP duration because there are multiple other more influential determinants. There are other non-schedule related strategies for reducing fatigue at TOD during (long) night FDPs (e.g. Wesensten et al. 2015; Gander, 2015; Dawson & McCulloch, 2005). One of them is strategic sleep before and during a flight. For example, in-flight sleep during long night flights with augmented flight crew has been found to be beneficial to fatigue at TOD (e.g. Van den Berg et al. 2016; Gander et al. 2013).

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

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

Early starts

Of all the disruptive duties the early starts turned out to be associated with the lowest fatigue scores at TOD. The analyses based on the entire data did not show significant findings, whereas the odds of high fatigue were doubled as compared to daytime FDPs in the between-subjects data. We also made an attempt to include a part of the night FDPs called deep early starts (start time between 02:00h and 04:59h) in the early starting FDPs. This attempt did not, however, yield results that would have markedly differed from the original ones. The fact that the evidence of increased fatigue at TOD was uncompelling can be explained by two factors: i) a relatively good ratio between the prior sleep and wake; largely because crew have been awake for a relatively short time at TOD at the end of an early duty; and ii) on early during the final TOD does not encroach on the WOCL. Also, our data did not cover the first two phases of a flight (blocks off and top of climb). It is possible that fatigue is actually higher during this part of an early start FDP because of the influence of the WOCL.

Late finishes

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

Nights

Our results demonstrate the need to further mitigate fatigue while flying during the night. When considering mitigation strategies, it is important to note that the present study did not reveal FDP characteristics (except for encroachment on the WOCL) that would have predicted high fatigue at TOD during night FDPs. In other words, fatigue at TOD was independent of the FDP-related characteristics, as long as an FDP fell into the night FDP category. Given this result, it is difficult to propose any scheduling-based solution to mitigate high fatigue at TOD during night FDPs.

Interestingly, our additional analyses revealed that especially deep night FDPs (end time after the WOCL) involved high fatigue at TOD. This finding suggests that the deep night FDP needs special attention when mitigating duty fatigue.

When interpreting our results of night FDPs, it is important to notice that high fatigue is to some extent an inevitable part of night work across industries because human beings are day-oriented (Åkerstedt, 1988; Monk, 1990; Sallinen & Hublin, 2015).

Need to revise regulations?

We presented an example of an alternative way to categorise FDPs typical of disruptive schedules. First, we re-categorised the night FDPs with a start time between 02:00h and 04:59h as ‘deep early starts’. A reason for this change was that crew were able to obtain some night sleep just before these very early start FDPs, unlike the other night FDPs. Another reason was to determine if deep early FDPs involved more fatigue than early FDPs (start time between 05:00h and 06:59h).

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

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

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

Chapter 4: Critique of the whole data collection activity

Critique assessment

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

Resulting critique of the data collection

Recruitment and training of crew members

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

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

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

Data collection tools

When crew members registered to collect data, they were asked whether they were willing to wear an actigraph. The actigraphs were sent out to the volunteers based on availability of the actigraphs. The fact that we could not send out actigraphs to all participants turned out to be confusing to some participants. Some volunteers explained that they were under the impression they had to wait until an actigraph was available. Others thought they had to choose between wearing the actigraph and collecting data with the app. This information about the confusion was based on several email conversations with participants in reply to sending a reminder for the FTL data collection.

Our communication indicated that the data collection could start without the actigraph. Although most volunteers understood this straight away, we had to explain the situation to volunteers who did not directly understand this, a number of times. Normally it was enough to show them the original email in order to make this clear.

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

References

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

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

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

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

Gander, P. H., Signal, T. L., van den Berg, M. J., Mulrine, H. M., Jay, S. M., & Jim Mangie, C. (2013). In-flight sleep, pilot fatigue and Psychomotor Vigilance Task performance on ultra-long range versus long range flights. *Journal of Sleep Research, 22(6)*, 697-706.

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

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

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

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

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

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

List of abbreviations

Abbreviations	*Description*
CI	Confidence Interval
D	Deliverable
EASA	European Aviation Safety Agency
EC	European Commission
EU	European Union
FDP	Flight Duty Period
FTL	Flight Time Limitation
KSS	Karolinska Sleepiness Scale
OR	Odds Ratio
PVT	Psychomotor Vigilance Task
TOD	Top Of Descent
WOCL	Window Of Circadian Low

European
Commission