Effectiveness of Flight Time Limitation (FTL)

Working Document 2.3 *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

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 working document

This Working Document 2.3 (Performance of the Data Collection and Data Analysis) reports the results of the work performed on the collection and analysis of the gathered data. 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: Data analysis methods

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 ordinal scales of the Karolinska Sleepiness Scale (KSS)[3] ≥ 7 and Samn-Perelli (SP)[4] ≥ 6. For a detailed description of the dependent (and independent) variables see Appendix 1: Variables list. Total sleep in 24 hours prior to TOD (Sleep24h) and Napping during 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 performance levels start to decrease at and above this level (e.g. Åkerstedt et al. 2014).

<u>FDP Baseline set</u>

This step determined the probabilities of the occurrence of high fatigue scores for an FDP Baseline set. The FDP Baseline set consists of all FDP data points available in the dataset. FDP Baseline probabilities were utilized in the secondary-objective analyses for FDP1 and FDP2 to determine ratios of the occurrence of high fatigue scores during the two FDPs of interest compared to the baseline.

<u>Primary objective FDP 1 (Night duties of more than 10 hours)</u>

The objective was to assess the prevalence of high fatigue scores during duties of more than 10 hours at the less favourable time of day (between 02:00h and 04:59h). To this end, the following two operational hypotheses were formulated:

H_0 = High fatigue scores do not occur in flight duties longer than 10 hours that take place between 02:00h and 04:59h.
H_1 = High fatigue scores occur in flight duties longer than 10 hours that take place between 02:00h and 04:59h.

---

[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.

The assessment consisted of estimating the probability that the KSS was scored 7 or higher, or the SP was scored 6 or higher. A point estimate as well as a confidence interval (CI) for the KSS and SP occurrence probabilities were determined[5].

Secondary objective FDP1 (Night duties of more than 10 hours)

The objective was to assess whether or not high fatigue scores during duties of more than 10 hours at the less favourable time of day (between 02:00h and 04:59h) occur more frequently than in the FDP Baseline set[6]. To this end, the following two operational and statistical hypotheses were formulated:

H_0 = High fatigue scores do not occur more frequently in flight duties longer than 10 hours that take place between 02:00h and 04:59h as in the FDP Baseline set.
H_1 = High fatigue scores occur more frequently in flight duties longer than 10 hours that take place between 02:00h and 04:59h than in the FDP Baseline set.

The assessment consisted of calculating the ratio for FDP1 compared to the FDP Baseline set. The ratio was defined as the ratio of the occurrence probabilities of individual or high fatigue scores for the two datasets.

Primary objective FDP2 (Disruptive schedules)

The objective was to assess the prevalence of high fatigue scores during consecutive disruptive FDPs. To this end, the following two operational hypotheses were formulated:

H_0 = High fatigue scores do not occur in consecutive disruptive FDPs, irrespective of number of repetitions and type: early start, late finish, night, or mix.
H_1 = High fatigue scores occur in consecutive disruptive FDPs, irrespective of number of repetitions and type: early start, late finish, night, or mix.

The FDP2 assessment was similar to the assessment of the primary objective for FDP1. The assessment focused firstly on the full FDP2 (Disruptive schedules) set containing all disruptive FDPs, irrespective of the number of repetitions, followed by the four types of consecutive disruptive duties (i.e., at least two in a row).

Secondary objective FDP2 (Disruptive schedules)

The objective was to assess whether or not high fatigue scores during consecutive disruptive FDPs occur more frequently than in the FDP Baseline set. To this end, the following two operational and statistical hypotheses were formulated:

H_0 = High fatigue scores do not occur more frequently in consecutive disruptive FDPs, irrespective of number of repetitions and type: early start, late finish, night, or mix, as in the FDP Baseline set.
H_1 = High fatigue scores occur more frequently in consecutive disruptive FDPs, irrespective of number of repetitions and type: early start, late finish, night, or mix, than in the FDP Baseline set.

The FDP2 assessment was similar to the assessment of the secondary objective for FDP1.

[5] Concerning the point estimate, two cases were distinguished, namely a point estimate equal to zero or a point estimate larger than zero. Zero observed occurrences lead to a formal point estimate of zero but do not imply that the true occurrence rate equals zero. The true occurrence rate (with the specified level of confidence) may be as large as the upper limit of the estimated CI. Non-zero observed occurrences result in a point estimate larger than zero. The corresponding CI may or may not include the value of zero. When the interval does not include zero, the occurrence probability of high levels of fatigue is significant.
[6] It should be emphasized that this secondary objective was introduced to describe the high fatigue observations in relation to the mean levels of fatigue in order to identify the high points in the data. No conclusions on effects or causes can and will be drawn from this.

Step 2: Compare fatigue scores between FDP categories

The differences between KSS scores in FDP1 (Night duties of more than 10 hours) and its control FDP categories and in FDP2 (Disruptive schedules) and its control FDP categories were assessed to gain more insight into which types of FDP categories are associated with fatigue. We used multiple logistic regression for this analysis.

For the assessment, the following operational and statistical hypotheses were formulated:

H_0 = The proportions of high fatigue in the FDP under consideration and a control group are not diferent.
H_1 = The proportions of high fatigue in the FDP under consideration and a control group are different.

The main approach for FDP1 (Night duties of more than 10 hours) was to compare the level of fatigue at TOD of short FDPs (≤ 10h) and long (> 10h) FDPs with adjustment for factors that may influence the outcome. A second criterion was that both long and short FDPs included at least one minute of the window of circadian low (WOCL).

Additional FDP1 comparisons were performed for cut-offs other than 10 hours:

- All night FDPs with a duration of 8 hours as cut-off (all ≤ 8 hours FDPs as short FDPs vs all > 8 hours FDPs as long FDPs);
- All night FDPs with a duration of 9 hours as cut-off (all ≤ 9 hours FDPs as short FDPs vs all > 9 hours FDPs as long FDPs);
- All night FDPs with a duration of 11 hours as cut-off (all ≤ 11 hours FDPs as short FDPs vs all > 11 hours FDPs as long FDPs); and
- All night FDPs with a duration of 12 hours as cut-off (all ≤ 12 hours FDPs as short FDPs vs all > 12 hours FDPs as long FDPs).

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

The statistical method used was logistic regression, since the interest was in determining the occurrence of high fatigue scores (KSS ≥ 7), rather than comparing mean values of fatigue (which would have involved a mixed model analysis of variance). Furthermore, the logistic regression was of the multiple type, always including a disruptive FDP type vs a non-disruptive FDP type, Function, Age, and Gender. Thereafter, each schedule- or sleep-related variable (predictor) was tried one at a time in the regression to investigate if it would result in a non-significant Odds Ratio (OR) for the disruptive/non-disruptive FDP variable.

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 can be found in Appendix 1: Variables list. These 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; and
- Ideas and suggestions from scientific committee and consortium members.

The variables that may contribute to the prediction of high levels of fatigue (i.e., KSS ≥ 7 and SP ≥ 6) were evaluated in a sequence of multiple logistic regression analyses. Performing a sequence of analyses allowed exploring the dynamics of the predictors. Logistic regression predicts the occurrence of a binary dependent variable, i.e. an event that either takes place (i.e., a high level of fatigue) or does not take place (i.e., a low level of fatigue). Logistic regression models the relationship between a binary dependent variable and one or more nominal, ordinal, interval or ratio-level independent variables. Multinomial logistic regression extends logistic regression to the case where a dependent variable can take more than two (discrete) values[7].

Step 4: Compare measurement tools

The different measurement tools that were used in the data collection were compared to determine whether or not the different tools provided the same results.

<u>Compare actigraphy and sleep log</u>

- Is there a significant correlation between sleep duration determined with actigraphy and sleep log?

<u>Compare measures KSS and SP</u>

- Are there significant correlations between self-ratings on the KSS and SP?

[7] For more information on logistic regression analyses check Hosmer et al. (2013).

Chapter 3: Data analysis results

Description of crew member data

D1 Addendum provides an overview of the candidate airlines to participate in data collection. The following 24 airlines accepted our invitation to participate in the field study:

Air Baltic, Air Europa, Alitalia, ASL Airlines Belgium, BRA Braathens, British Airways, Cargolux, Condor, Czech Airlines, Flybe, Iberia, KLM, Lufthansa, Lufthansa Cargo, Norwegian Air Int., Norwegian Air Shuttle, Scandinavian Airlines, TAP Portugal, TAROM, Vueling, WIZZ Air, SunExpress Deutschland[8], Adria Airways, and Croatia Airlines.

Four geographical European regions were defined in D1 Addendum: East, West, North, and South. Table 1 shows the number of participating airlines per region.

Table 1 No. of airlines participating across Europe

	East	*West*	*North*	*South*
No. of airlines	6	9	4	5

Five types of operation were defined in D1 Addendum. Table 2 provides an overview with regard to the types of operation of the participating airlines. The total number of airlines exceeds 24 as some airlines operate more than one type of operation.

Table 2 Types of operation of the participating airlines

	Long-haul	*Medium-haul*	*Short-haul*	*Regional*	*Cargo*
No. of airlines	11	12	18	10	3

Volunteers registered for participation in the data collection via an online portal. After registering, giving consent to participate, and familiarising themselves with the data collection application (app) and measurement protocol, the volunteers started their 14 days of data collection.

Figure 1 shows the flow of the initial number of registrations, followed by drop-outs and insufficient quality of collected data to the resulting number of crew members proving adequate quality data and FDPs. A total of 2877 FDPs were gathered.

Crew members needed to be acclimatised[9] for the FDPs to be included in the dataset. A crew member is considered to be acclimatised to a 2-hour wide time zone surrounding the local time at the point of departure. If the crew was not acclimatised during any part of the FDP, the corresponding duty was excluded from the dataset. This resulted in 173 excluded FDPs.

A further four FDPs were excluded because they were performed by an augmented crew.

[8] SunExpress Deutschland volunteered without being invited explicitly by their National Aviation Authority.
[9] According to ORO.FTL.105, acclimatised is defined as a state in which a crew member's circadian biological clock is synchronised to the time zone where the crew member is.

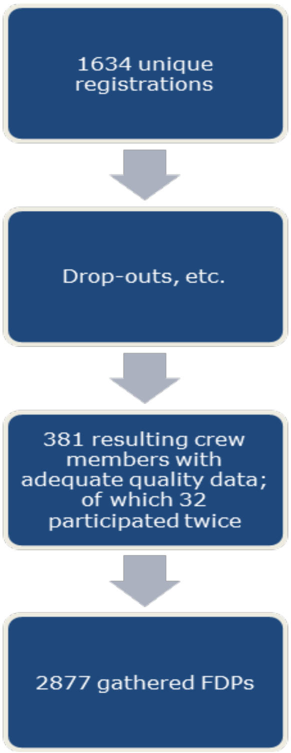


Figure 1 Flow of registration to gathered FDPs[10]

Table 3 and Table 4 provide information on the crew member demographics and type of operators that collected data.

Table 3 Crew member demographics

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

M: Male. F: Female. U: Unknown. SD: Standard Deviation. Min: minutes. Yr: years of age.

Table 4 Type of operator[11]

	Participants: *No. (%)*	*Pilot/Cabin*
Network operator	173 (45%)	107/66
Point-to-point operator	130 (34%)	78/52
Cargo operator	78 (21%)	76/2
Total	**381 (100%)**[12]	**261/120**

[10] Thirty-two participants (of the 381) volunteered to perform the 14-days of data collection twice (in two separate periods, not in one stretch).
[11] Participants recorded the type of operator they work for on the data collection app. The numbers in Table 4 refer to this recording.

A total of 2877 FDPs (i.e., the FDP Baseline set) were gathered between July 2017 and February 2018 (Figure 2).

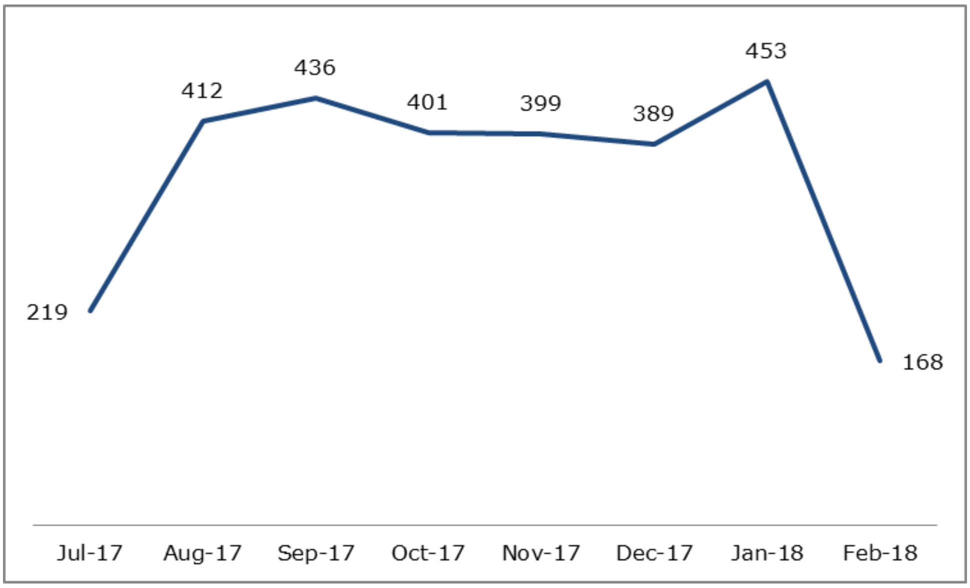


Figure 2 No. of gathered FDPs per month (2017 - 2018)

The following tables show the number of gathered FDP1 (Night duties of more than 10 hours), FDP2 (Disruptive schedules), and FDP Baseline set in relation to the required sample size as estimated in D2.2 (Definition of the Data Collection Process).

Table 5 Sample sizes for FDP1 (Night duties > 10h), FDP2 (Disruptive schedules), and FDP Baseline set

	FDP1 (Night duties > 10h) *No.*	***Full FDP2 (Disruptive schedules) set*** *No.*	***FDP Baseline*** *No.*
Pilot FDPs	136	822	1932
Cabin FDPs	65	354	945
Estimated required total sample size	*134*	*Not applicable**	*Not applicable***
Total FDP sample size	**201**	**1176**	**2877**

* The estimated sample size for FDP2 in D2.2 was based on consecutive Disruptive schedules whereas the table refers to non-consecutive Disruptive schedules.
** No sample sizes were calculated in D2.2 for FDP Baseline set.

Table 6 Sample sizes for FDP2 (Disruptive schedules): Early starts, Late finishes, and Nights

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

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

Table 7 Sample sizes for consecutive FDP2 (Disruptive schedules) (at least two FDPs in a row): Consecutive early starts, late finishes, nights, and mix

	Consecutive early starts *No.*	***Consecutive late finishes*** *No.*	***Consecutive nights*** *No.*	***Consecutive mix*** *No.*
Pilot FDPs	49	11	94	87
Cabin FDPs	31	12	29	26
Estimated required total sample size	*108*	*108*	*108*	*108*
Total FDP sample size	**80**	**23**	**123**	**113**

Table 8 Sample sizes for consecutive FDP2 (Disruptive schedules) (two or more FDPs in a row): Consecutive early starts, late finishes, and nights

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

Table 9 Sample size FDP2 (Disruptive schedules): Mix

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

Results for FDP1 (Night duties of more than 10 hours)

Check for high fatigue scores

This section begins with the primary objective for the individual duty period FDP1 (Night duties of more than 10 hours) by presenting point estimates and 95% CI for the occurrence probabilities of individual outcome measure values as well as for high levels of fatigue. Following that, the same results are presented for the FDP Baseline set. Subsequently, the secondary objective is addressed by examining the ratio for FDP1 compared to the FDP Baseline set.

Results are presented for KSS TOD and SP TOD. Other than missing values, the outcome measures were not affected by outliers. All computations were performed in Microsoft Excel.

Primary objective for FDP1 (Night duties of more than 10 hours)
The objective was to assess the prevalence of high fatigue scores during duties of more than 10 hours at the less favourable time of day (between 02:00h and 04:59h).

The operational hypotheses were addressed by estimating the probability of the occurrence of high fatigue scores where high fatigue was characterized by KSS scores higher than 6 or SP scores higher than 5. In fact, occurrence probabilities were also estimated for the individual KSS and SP scores.

Table 10 presents occurrence-probability point estimates[13] as well as 95% lower and upper confidence limits for KSS and SP. The number of occurrences of the high fatigue scores and the number of valid measurements (Nvalid) are included. As can be seen from the estimates, high fatigue scores did occur in the flight duties longer than 10 hours that encroach (part of) the period between 02:00h and 04:59h as postulated under the hypothesis H_1.

Table 10A KSS TOD and SP TOD occurrence-probability point estimates for FDP1 (Night duties > 10h)

Fatigue measure	N	*p*	*p*L	*p*U
KSS TOD 7, 8, 9	60	0.414	0.337	0.495
Nvalid	145			
SP TOD 6, 7	28	0.196	0.139	0.268
Nvalid	143			

Occurrence-probability point estimates (*p*) as well as 95% lower (*p*L) and upper confidence limits (*p*U).

Table 10B KSS TOD and SP TOD occurrence-probability point estimates for FDP1 (Night duties < 10h)

Fatigue measure	N	*p*	*p*L	*p*U
KSS TOD 7, 8, 9	106	0.302	0.256	0.352
Nvalid	349			
SP TOD 6, 7	43	0.123	0.093	0.161
Nvalid	351			

Occurrence-probability point estimates (*p*) as well as 95% lower (*p*L) and upper confidence limits (*p*U).

[13] Note that with, regard to hypothesis H_0, not observing any high fatigue scores in a dataset does not imply that the underlying occurrence probability was zero. If the observed number of high fatigue scores was zero, then we used the 95% upper confidence limit as a conservative point estimate.

The table below presents the descriptive statistics of the total sleep in 24 hours prior to TOD (Sleep24h) and Napping during FDP (FDPsleep) for the high (KSS ≥ 7) and low levels of fatigue (KSS < 7). The probability of sleep *not* to occur in 24 hours prior TOD is 0.174 for FDP1 (Night duties of more than 10 hours). The probability of a nap to occur during a duty is 0.274 for FDP1.

Table 11 Sleep24h and FDPsleep for high and low KSS TOD scores for FDP1 (Night duties > 10h)

		Min	*Max*	*Mean*	*SD*
Sleep24h	KSS High	15	385	100.95	108.876
(in min)	KSS Low	10	485	232.60	118.709
FDPsleep	KSS High	10	30	17.67	6.155
(in min)	KSS Low	5	30	18.96	6.830

SD: Standard Deviation.

The figure below presents the percentages of high and low fatigue scores (KSS) for each hour of the day.

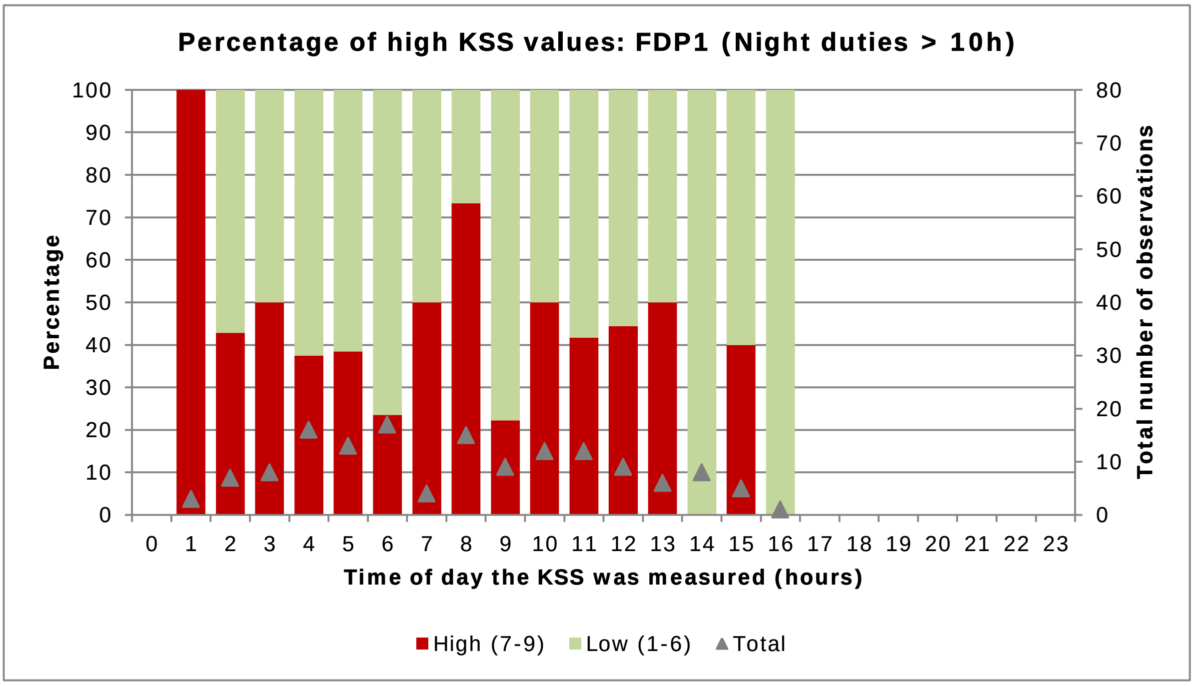


Figure 3 Percentages of high and low KSS TOD scores for each hour of the day for FDP1 (Night duties > 10h)

The KSS and SP measures differed with regard to their estimates of the occurrence probabilities of high levels of fatigue, where the KSS measure used three values for high fatigue and the SP measure two. To examine the effect, if any, of the broader range of high level of fatigue values for the KSS measure, an occurrence probability of a high level of fatigue was computed for KSS = 8 and KSS = 9 only. Table 12 summarizes the (point) estimates of the high fatigue scores for the different cases. The table shows that restricting the range of high scores for the KSS measure to 8 and 9 resulted in a much stronger similarity between the KSS and SP probabilities.

Table 12A Point estimates of the high fatigue scores for FDP1 (Night duties > 10h)

KSS TOD		***SP TOD***
7, 8, 9	**8, 9**	**6, 7**
0.414	0.214	0.196

Table 12B Point estimates of the high fatigue scores for FDP1 (Night duties < 10h)

KSS TOD		***SP TOD***
7, 8, 9	**8, 9**	**6, 7**
0.302	0.172	0.123

FDP Baseline set

The FDP Baseline set consists of all FDP data points available in the dataset. There were no particular objectives for the FDP Baseline set other than it being used for an assessment of the ratio for FDP1 (as well as for FDP2). The occurrence probabilities of the different fatigue measure values of the FDP Baseline set, together with their 95% upper and lower confidence limits, are presented in a similar manner as for FDP1.

Table 13 presents the occurrence-probability point estimates as well as 95% lower and upper confidence limits for KSS TOD and SP TOD. High fatigue scores did occur in the FDP Baseline set of all gathered FDPs.

Table 13 KSS TOD and SP TOD occurrence-probability point estimates for the FDP Baseline set

Fatigue measure	N	*p*	*p*L	*p*U
KSS TOD 7, 8, 9	423	0.241	0.222	0.262
Nvalid	1755			
SP TOD 6, 7	162	0.092	0.080	0.107
Nvalid	1757			

Occurrence-probability point estimates (*p*) as well as 95% lower (*p*L) and upper confidence limits (*p*U).

For the FDP Baseline set, the size of the 95% CI was smaller than for FDP1 due to the larger sample size.

The KSS and SP measures differed with regard to their estimates of the occurrence probabilities of high levels of fatigue, where the KSS measure used three values for high fatigue and the SP measure two. To examine the effect, if any, of the broader range of high level of fatigue values for the KSS measure, an occurrence probability of a high level of fatigue was also computed for KSS = 8 and KSS = 9 only. Table 14 summarizes the (point) estimates of the high fatigue scores for the different cases. The table shows that restricting the range of high fatigue values for the KSS measure to 8 and 9 resulted in a much greater similarity between the KSS and SP high fatigue probabilities.

Table 14 Point estimates of the high fatigue scores for the FDP Baseline

KSS TOD		***SP TOD***
7, 8, 9	**8, 9**	**6, 7**
0.241	0.121	0.092

Secondary objective for FDP1 (Night duties of more than 10 hours)

The objective was to assess whether or not high fatigue scores during duties of more than 10 hours at the less favourable time of day (between 02:00h and 04:59h) occur more frequently than in the FDP Baseline set.

The ratio estimates for FDP1 (Night duties of more than 10 hours) and the FDP Baseline set for the high KSS and SP scores were 1.717 and 2.124. Both estimates were considerably larger than 1.0, showing that FDP1 was more prone to high fatigue scores than the FDP Baseline set.

It has also been assessed whether or not high fatigue scores during short-night FDPs occur more frequently than in the FDP Baseline set.

The ratio estimates for short-night FDPs and the FDP Baseline set for the high KSS and SP scores were 1.046 and 0.971. These estimates are marginally larger and smaller than 1.0, respectively, suggesting that short-night FDPs are similarly prone to high fatigue scores as the FDP Baseline set.

INTERIM CONCLUSION

The goal of this step was to assess the prevalence of high fatigue scores during flight duties longer than 10 hours that encroach (part of) the period between 02:00h and 04:59h.

What we have learned from this step is that high fatigue scores do occur and that the proportion of high fatigue is higher than in the FDP Baseline set containing all collected FDPs.

Note that this baseline contains all types of flights, including long-night FDPs, and that we do not know whether this baseline is indeed representative, for example in terms of distribution of flight duties across different FDP categories.

Compare fatigue scores between long- and short-night FDPs

The main approach of the following analyses was to compare the condition of interest with a control condition; i.e., long-night FDPs with short-night FDPs. The analyses involved logistic regression, that is, the outcome variable is a high or low level of fatigue (coded 1 and 0, respectively). However, in order to illustrate the results on the KSS, we also computed the means and standard deviation of the different FDPs.

Figure 4 shows the mean KSS ratings at TOD for the long- and short-night FDPs separately. For this between-subject analysis, participants were assigned to the long- and/or short-duration groups as follows. For participants with one or more FDPs in the short-duration group and none in the long-duration group, the participants' first FDP (and that one only in case of multiple FDPs) was taken for the short-duration group. Similarly, for participants with one or more FDPs in the long-duration group and none in the short-duration group, the participants' first FDP (and that one only in case of multiple FDPs) was taken for the long-duration group. Finally, for participants with at least one FDP in each of the groups, the participants' first FDPs meeting the respective groups' criteria were taken. Thus, the comparison of short with long FDPs contains different participants in the respective groups.

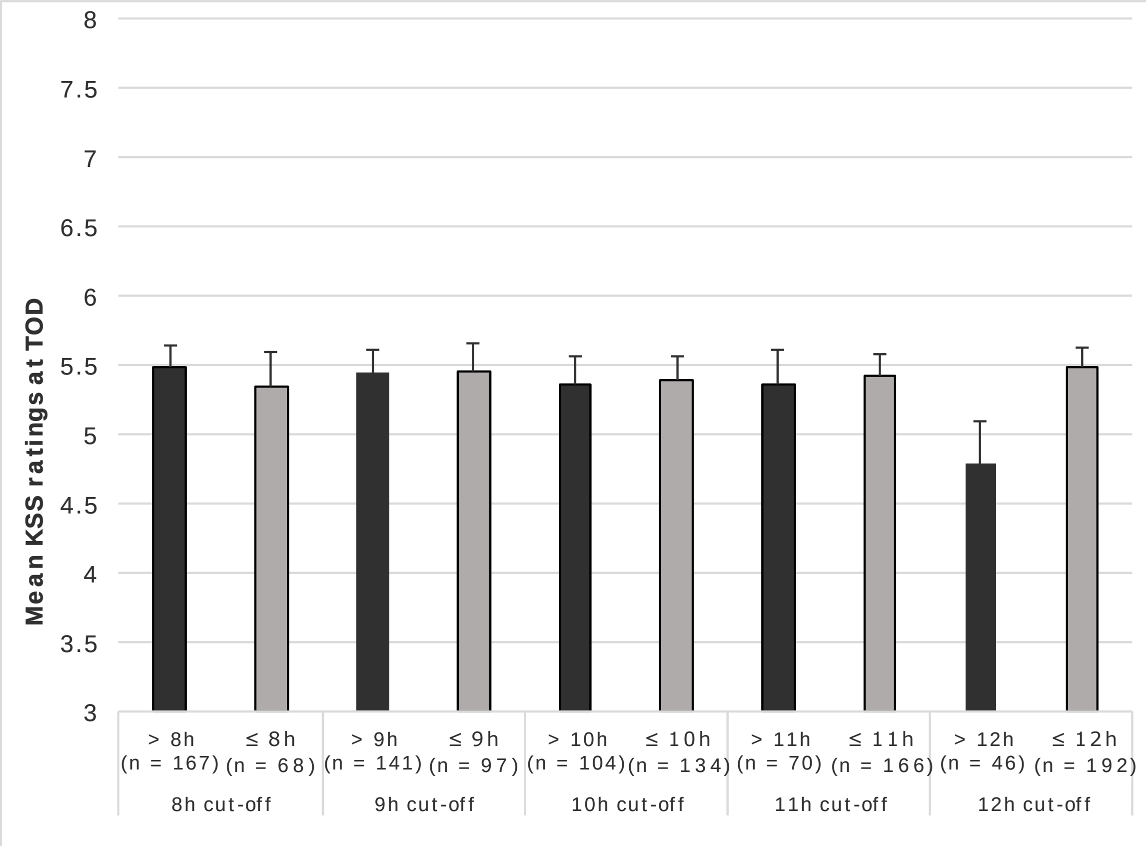


Figure 4 Mean KSS ratings at TOD for the *long*- and *short-night* FDPs. The criterion for the duration of a long-night FDP ranged from > 8h to > 12h. The black and grey bars indicate the long and short FDPs in the between-subjects data. The vertical lines denote the standard errors. Note that the y-axis covers only part of the 9-point scale

The mean KSS ratings ranged between 5.0 and 5.5 in all long- and short-night FDP categories, except for the FDPs longer than 12 hours in the between-subjects data (mean KSS = 4.8, Standard Error SE = 0.3). These descriptive statistics showed no systematic differences between the long- and short-night FDPs across the comparisons where the criterion for the duration of a long-night FDP varied between > 8 hour and > 12 hour. Note that no statistical tests or adjustments for other factors were applied.

The subsequent multiple logistic regression analysis used the between-subject FDPs as set out above. Thus, the comparison of long with short FDPs contains different participants in the respective groups. We used a multiple logistic regression with the level of fatigue as dependent variable. High levels of fatigue, KSS ≥ 7, were assigned the value of “1”. The reference was low levels of fatigue, KSS < 7, which were assigned the value of “0”. In order to account for differences between pilots (1) and cabin crew (2), this variable was entered in all analyses. Furthermore, all analyses were adjusted for Age (in years) and Gender (1 = male, 2 = female).

The major statistic for outcome in these analyses was the Odds Ratio (OR) and its 95% CI. The OR indicates how many times larger the odds are for the exposed group to report high levels of fatigue than the odds in the reference group (in this case, short-night FDPs). The result is significant if the CI does not overlap “1”. An OR of 1.5 indicates that the exposed group is 50% more likely to report high levels of fatigue than the reference group. An OR = 0.5 means that the exposed group is 50% less likely to report high levels of fatigue than the reference group. Note that some predictors have only two categories (e.g. long/short), while others have many (e.g. sleep, expressed in hours). The OR value is computed per category of the predictor. Thus ORs for a variable with two categories are not comparable in size to a variable with many categories.

In order to adjust for other variables that may affect the OR of the main analysis, each such variable was entered one at a time as a covariate. If the analysis showed that the covariate reduced a significant OR into a non-significant value, we considered this variable as having had an effect on the OR. Only significant covariates are included in the tables, that is, those not listed did not affect the OR of the FDP category.

The variables included were those thought to be relevant for regulation or strongly affected by scheduling (like sleep and time awake):

- FDP duration (FDPdur, in minutes);
- Time zones crossed from East to West (TZ_EW, in numbers);
- Time zones crossed from West to East (TZ_WE, in numbers);
- Time in WOCL (FDPWOCLm, in minutes);
- KSS TOD rated in/out WOCL (WOCLTOD, 0 = outside WOCL, 1 = inside WOCL);
- Sleep during FDP (FDPsleep, in minutes);
- Time of day KSS TOD (TimeDayTOD, in hours);
- Time awake since last sleep (AwakeTOD, in minutes);
- Sleep in 24 hours prior TOD (Sleep24h, in minutes);
- Sectors flown in FDP (FDPsectors, in numbers);
- FDP start time (StartH, in hours); and
- FDP end time (EndH, in hours).

Variables that involved considerable data loss (> 30% loss) were excluded: Total FDP duration in one week (FDPdurWk, in min) and Time off prior to FDP (RestPeriod, in min). These variables also had weak correlations with fatigue ($r < .05$).

Table 15 shows that the OR for long-/short-night FDPs (using the 10 hours cut-off) is close to 1 with a CI that overlaps 1. This non-significant result makes adjustment for variables less meaningful, although variables could make the main association significant. The OR for Function was not significant, although there seems to be a trend of higher likelihood for cabin crew to report high levels of fatigue. None of the other variables had any effect on the ORs when entered one at a time.

Table 15 Results from logistic regression predicting high levels of fatigue from long/short FDPs with cut-off at 10 hour, and from Function. N = 229

Predictors	***OR***	***95% CI***	***p***	***% high fatigue***
Long/short (1/0)	1.060	0.608; 1.846	.818	36/33
Pilot/cabin (1/2)	1.900	0.948; 3.807	.070	

OR = Odds Ratio. CI = Confidence Interval. p = level of significance.
The result is adjusted for Age and Gender. % high fatigue = % with KSS ≥ 7 in the long-/short-night FDPs.

The percent with high levels of fatigue was very similar in the long- and short-night FDPs. Note that the percentage was high for both conditions, probably because both were night FDPs encroaching on the WOCL. When short/long FDPs were defined using 7, 8, 9, and 11 hours as cut-offs, no significant OR was obtained (not presented).

Finally, we computed the mean and Anova F-ratios for the two FDP categories for all the other variables. This was done in order to illustrate in what ways the two FDP categories differed on the other predictors of high levels of fatigue. Table 16 shows that long- and short-night FDPs differed on many of the analysed variables. Thus, longer FDPs had more time in the WOCL, had more time awake, longer FDP duration (for obvious reasons), less prior sleep, more time zones crossed in either direction, and ended later, than the short FDPs. The number of sectors did not differ. Thus, several long-night FDPS are characterized by several fatigue inducing factors, which would make one expect high fatigue. However, long-night FDPs also end later than short ones, that is, they end at times when the circadian rhythm should have increased alertness considerably compared to the short-night FDPs, which ended between 06:00h and 06:59h, close to the probable circadian trough (low point) of alertness. We assume that this difference in end times at least partly contributes to the lack of difference between long and short night flights.

Table 16 Anova results, mean±SD for long- and short-night FDPs (10h cut-off). No adjustment. N = 238

	Short *Mean±SD*	***Long*** *Mean±SD*	***F-ratio***	***p***
FDPWOCLm: Time in WOCL	2.25±1.15	3.39±1.06	59.57	.000
FDPsectors: Sectors flown	1.76±0.91	1.67±1.01	0.50	.478
AwakeTOD: Time awake prior TOD	12.79±6.65	20.12±6.76	68.79	.000
EndH: FDP end time	6.67±3.16	9.18±3.98	29.45	.000
FDPdur: FDP duration	6.67±3.16	11.82±1.40	390.37	.000
Sleep24h: Sleep in 24h prior TOD	5.15±2.51	3.28±2.64	30.28	.000
TimeDayTOD: Time of day at TOD	5.74±3.19	8.25±3.78	28.88	.000
StartH: FDP start time	00.11±3.69	21.36±3.77	32.08	.000
FDPsleep: Napping during FDP	2.67±7.32	5.15±9.39	4.10	.045
TZ_EW: Time zones crossed EW	0.47±1.22	1.60±2.56	17.77	.000
TZ_WE: Time zones crossed WE	1.53±2.62	3.15±3.35	15.17	.000

SD = Standard Deviation. p = level of significance.

We also carried out the analysis presented in Table 15 using KSS MAX ratings. The latter was obtained from other legs during the FDP and/or cruise phase during long-haul flights. Using these ratings, we took the highest one (including those at TOD) to represent maximum fatigue. In the present analysis the result was very similar to that in Table 15 – no significant difference between long- and short-night FDPs.

INTERIM CONCLUSION

The goal of this step was to compare short-night FDPs with long-night FDPs, with respect to the occurrence of high fatigue at TOD, and with adjustment for factors that may influence the outcome.

We created different FDP categories (with a main focus on the 10 hours cut-off between categories) and our interim conclusion is that short and long FDPs that encroach on the WOCL (night FDPs) do not differ in their odds of high fatigue at TOD.

Results for FDP2 (Disruptive schedules)

Check for high fatigue scores

This section begins with the primary objective for the individual duty period FDP2 (Disruptive schedules) by presenting point estimates and 95% CI for the occurrence probabilities of individual fatigue measure values as well as for high levels of fatigue. Following that, the same quantities are considered for the FDP Baseline set. Subsequently, the secondary objective is addressed by examining the ratio for FDP2 (Disruptive schedules) compared to the FDP Baseline set.

Results are presented for KSS TOD and SP TOD. Other than missing values, the four outcome measures were not affected by outliers. All computations were performed in Microsoft Excel.

For FDP2 (Disruptive schedules), the following subsets were defined in addition to the full FDP2 set:

- FDP2 (Early starts);
- FDP2 (Late finishes);
- FDP2 (Nights);
- Consecutive (i.e., at least two in a row) FDP2 (Early starts);
- Consecutive FDP2 (Late finishes);
- Consecutive FDP2 (Nights); and
- FDP2 (Mix).

These subsets for disruptive schedules are presented after the full FDP2 set. The FDP2 assessments were similar to the assessment of the primary objective for FDP1 (Night duties of more than 10 hours).

<u>Primary objective for the full FDP2 (Disruptive schedules) set</u>
The objective was to assess the prevalence of high levels of fatigue during disruptive schedule duties, irrespective of number of repetitions and type: early start, late finish, or night.

Table 17 presents occurrence-probability point estimates as well as 95% lower and upper confidence limits for KSS and SP. As can be seen from these estimates, high fatigue scores did occur in disruptive schedules as postulated under the hypothesis H_1.

Table 17 KSS TOD and SP TOD occurrence-probability point estimates for the full FDP2 (Disruptive schedules) set

Fatigue measure	N	*p*	*p*L	*p*U
KSS TOD 7, 8, 9	237	0.303	0.272	0.336
Nvalid	783			
SP TOD 6, 7	100	0.128	0.106	0.153
Nvalid	783			

Occurrence-probability point estimates (*p*) as well as 95% lower (*p*L) and upper confidence limits (*p*U).

The table below presents the descriptive statistics of the total sleep in 24 hours prior to TOD (Sleep24h) and Napping during FDP (FDPsleep) for the high (KSS ≥ 7) and low levels of fatigue (KSS < 7). The probability of sleep *not* to occur in 24 hours prior TOD is 0.080 for the full FDP2 (Disruptive schedules) set. The probability of a nap to occur during a duty is 0.122 for the full FDP2 (Disruptive schedules) set.

Table 18 Sleep24h and FDPsleep for high and low KSS TOD scores for the full FDP2 (Disruptive schedules) set

		Min	*Max*	*Mean*	*SD*
Sleep24h	KSS High	15	620	309.52	136.857
(in min)	KSS Low	10	720	344.39	135.337
FDPsleep	KSS High	3	42	19.54	8.709
(in min)	KSS Low	5	45	19.01	7.718

SD: Standard Deviation.

The figure below presents the percentages of high and low fatigue scores (KSS) for each hour of the day.

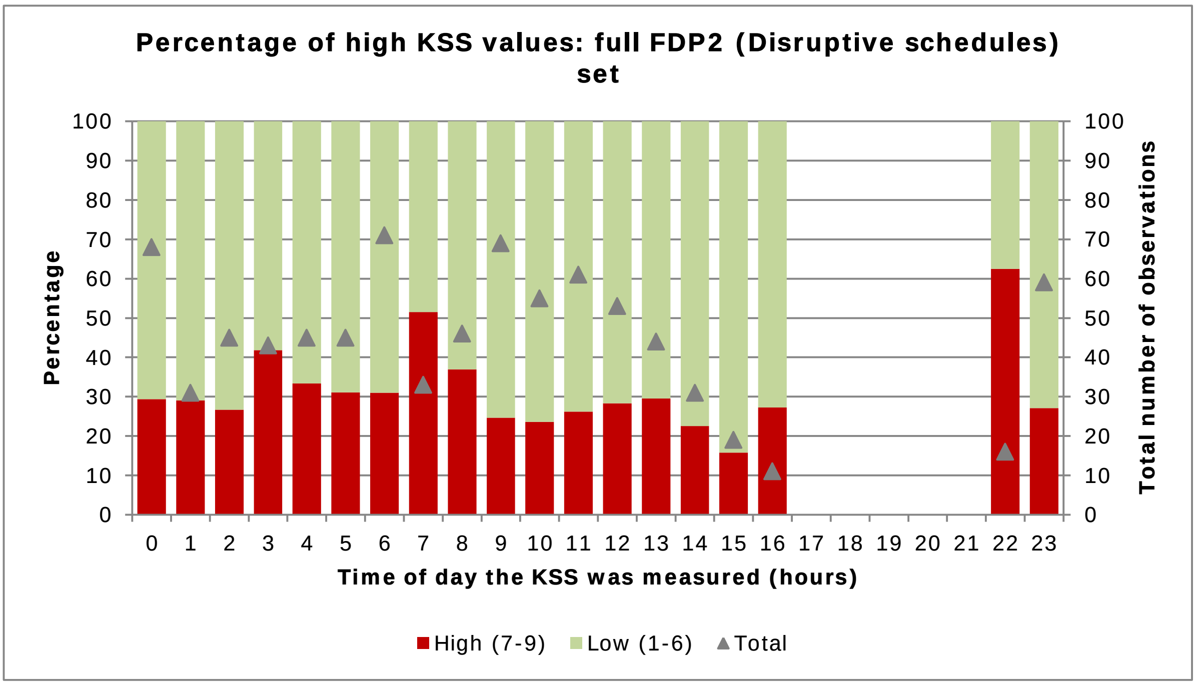


Figure 5 Percentages of high and low KSS TOD scores for each hour of the day for the full FDP2 (Disruptive schedules) set

The KSS scores differed from those of the SP. The two measures also differed with regard to their estimates of the occurrence probabilities of high levels of fatigue, where the KSS measure uses three values for high fatigue and the SP measure two. To examine the effect, if any, of the broader range of high level of fatigue values for the KSS measure, an occurrence probability of a high level of fatigue was also computed for KSS = 8 and KSS = 9 only. Table 19 summarizes the (point) estimates of the high fatigue scores for the different cases. The table shows that restricting the range of high fatigue values for the KSS measure to 8 and 9 only resulted in a much stronger similarity between the KSS and SP probabilities.

Table 19 Point estimates of the high fatigue scores for the full FDP2 (Disruptive schedules) set

KSS TOD		*SP TOD*
7, 8, 9	**8, 9**	**6, 7**
0.303	0.166	0.128

FDP Baseline set

An FDP Baseline was defined consisting of all FDP data points available in the dataset.

Results on the FDP Baseline set were already presented for FDP1 (Night duties of more than 10 hours).

Secondary objective for the full FDP2 (Disruptive schedules) set

The objective was to assess whether or not high fatigue scores during disruptive flight schedules occur more frequently than in the FDP Baseline set.

The ratio estimates for the full FDP2 (Disruptive schedules) set and the FDP Baseline set for the high KSS and SP scores were 1.257 and 1.391. Both estimates were considerably larger than 1.0, showing that the full FDP2 set was more prone to high fatigue scores than the FDP Baseline set.

Primary objective for FDP2 (Early starts)

The objective was to assess the prevalence of high levels of fatigue during early starts.

Table 20 presents occurrence-probability point estimates as well as 95% lower and upper confidence limits for KSS and SP. As can be seen from these estimates, high fatigue scores did occur in the early starts as postulated under the hypothesis H_1.

Table 20 KSS TOD and SP TOD occurrence-probability point estimates for FDP2 (Early starts)

Fatigue measure	N	*p*	*p*L	*p*U
KSS TOD 7, 8, 9	31	0.190	0.137	0.257
Nvalid	163			
SP TOD 6, 7	10	0.062	0.034	0.110
Nvalid	162			

Occurrence-probability point estimates (*p*) as well as 95% lower (*p*L) and upper confidence limits (*p*U).

The table below presents the descriptive statistics of the total sleep in 24 hours prior to TOD (Sleep24h) and Napping during FDP (FDPsleep) for the high (KSS ≥ 7) and low levels of fatigue (KSS < 7). The probability of sleep *not* to occur in 24 hours prior TOD is 0.032 for FDP2 (Early starts). The probability of a nap to occur during a duty is 0.054 for FDP2 (Early starts).

Table 21 Sleep24h and FDPsleep for high and low KSS TOD scores for FDP2 (Early starts)

		Min	***Max***	***Mean***	***SD***
Sleep24h	KSS High	110	620	351.85	101.953
(in min)	KSS Low	30	570	340.70	100.231
FDPsleep	KSS High	15	30	20.00	8.660
(in min)	KSS Low	6	25	17.75	7.225

SD: Standard Deviation.

The figure below presents the percentages of high and low fatigue scores (KSS) for each hour of the day.

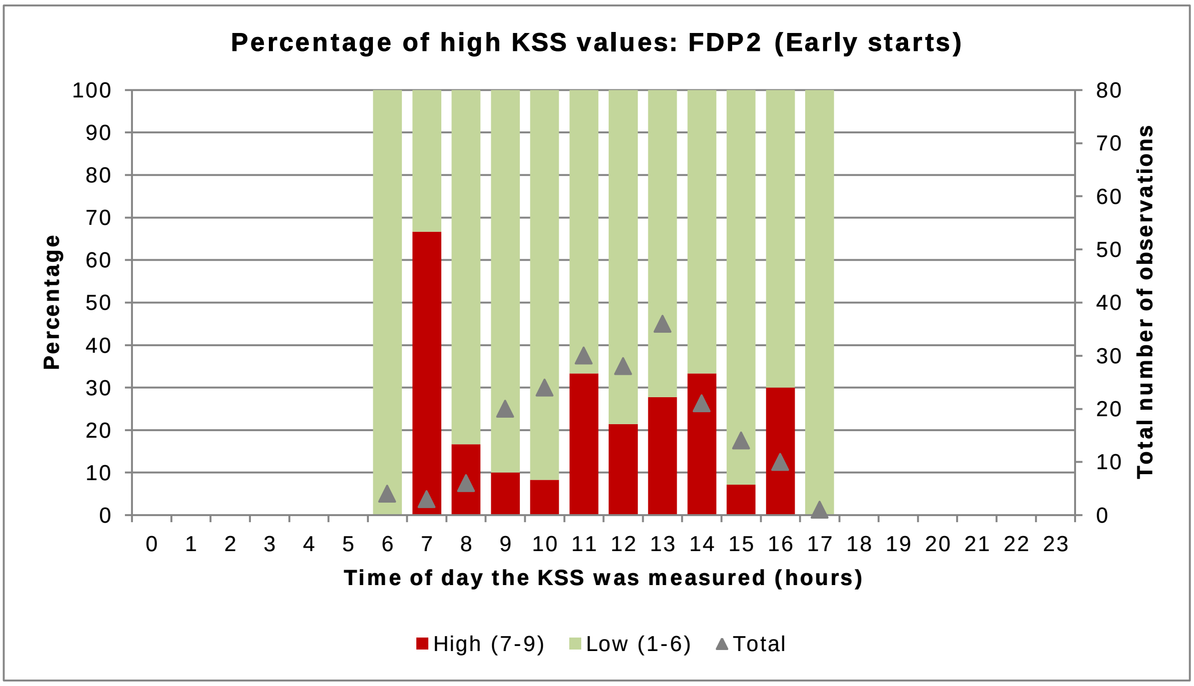


Figure 6 Percentages of high and low KSS TOD scores for each hour of the day for FDP2 (Early starts)

<u>Secondary objective for FDP2 (Early starts)</u>
The objective was to assess whether or not high fatigue scores during FDP2 (Early starts) occur more frequently than in the FDP Baseline set.

The ratio estimates for the FDP2 (Early starts) and the FDP Baseline set for the high KSS and SP scores were 0.788 and 0.674. Both estimates were smaller than 1.0. This showed that high fatigue scores did not occur more frequently in FDP2 (Early starts) than in the FDP Baseline dataset.

<u>Primary objective for FDP2 (Late finishes)</u>
The objective was to assess the prevalence of high levels of fatigue during late finishes.

Table 22 presents occurrence-probability point estimates as well as 95% lower and upper confidence limits for KSS and SP. As can be seen from these estimates, high fatigue scores did occur in the late finishes as postulated under the hypothesis H_1.

Table 22 KSS TOD and SP TOD occurrence-probability point estimates for FDP2 (Late finishes)

Fatigue measure	N	*p*	*p*L	*p*U
KSS TOD 7, 8, 9	38	0.309	0.234	0.395
Nvalid	123			
SP TOD 6, 7	17	0.137	0.087	0.209
Nvalid	124			

Occurrence-probability point estimates (*p*) as well as 95% lower (*p*L) and upper confidence limits (*p*U).

The table below presents the descriptive statistics of the total sleep in 24 hours prior to TOD (Sleep24h) and Napping during FDP (FDPsleep) for the high (KSS ≥ 7) and low levels of fatigue (KSS < 7). The probability of sleep *not* to occur in 24 hours prior TOD is 0.059 for FDP2 (Late finishes). The probability of a nap to occur during a duty is 0.108 for FDP2 (Late finishes).

Table 23 Sleep24h and FDPsleep for high and low KSS TOD scores for FDP2 (Late finishes)

		Min	*Max*	*Mean*	*SD*
Sleep24h	KSS High	240	600	421.74	83.716
(in min)	KSS Low	100	711	418.15	110.640
FDPsleep	KSS High	3	29	20.20	10.257
(in min)	KSS Low	10	30	18.08	6.825

SD: Standard Deviation.

The figure below presents the percentages of high and low fatigue scores (KSS) for each hour of the day.

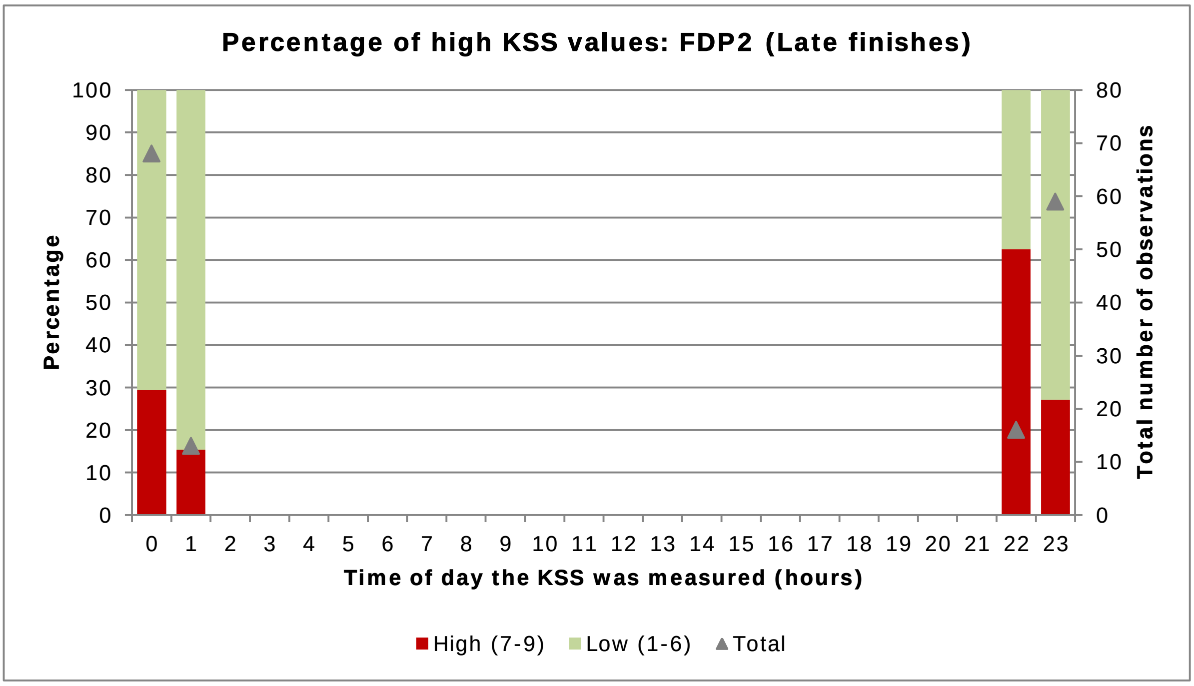


Figure 7 Percentages of high and low KSS TOD scores for each hour of the day for FDP2 (Late finishes)

<u>Secondary objective for FDP2 (Late finishes)</u>
The objective was to assess whether or not high fatigue scores during late finishes occur more frequently than in the FDP Baseline set.

The ratio estimates for the FDP2 (Late finishes) and the FDP Baseline set for the high KSS and SP scores were 1.282 and 1.489. Both estimates were larger than 1.0. This showed that high fatigue scores did occur more frequently in FDP2 (Late finishes) than in the FDP Baseline dataset.

<u>Primary objective for FDP2 (Nights)</u>
The objective was to assess the prevalence of high levels of fatigue during nights.

Table 24 presents occurrence-probability point estimates as well as 95% lower and upper confidence limits for KSS and SP. As can be seen from these estimates, high fatigue scores did occur in the nights as postulated under the hypothesis H_1.

Table 24 KSS TOD and SP TOD occurrence-probability point estimates for FDP2 (Nights)

Fatigue measure	N	*p*	*p*L	*p*U
KSS TOD 7, 8, 9	166	0.336	0.296	0.379
Nvalid	494			
SP TOD 6, 7	71	0.144	0.116	0.177
Nvalid	494			

Occurrence-probability point estimates (*p*) as well as 95% lower (*p*L) and upper confidence limits (*p*U).

The table below presents the descriptive statistics of the total sleep in 24 hours prior to TOD (Sleep24h) and Napping during FDP (FDPsleep) for the high (KSS ≥ 7) and low levels of fatigue (KSS < 7). The probability of sleep *not* to occur in 24 hours prior TOD is 0.106 for FDP2 (Nights). The probability of a nap to occur during a duty is 0.154 for FDP2 (Nights).

Table 25 Sleep24h and FDPsleep for high and low KSS TOD scores for FDP2 (Nights)

		Min	***Max***	***Mean***	***SD***
Sleep24h	KSS High	15	600	264.44	135.277
(in min)	KSS Low	10	720	321.28	149.427
FDPsleep	KSS High	5	42	19.37	8.780
(in min)	KSS Low	5	45	19.53	8.097

SD: Standard Deviation.

The figure below presents the percentages of high and low fatigue scores (KSS) for each hour of the day.

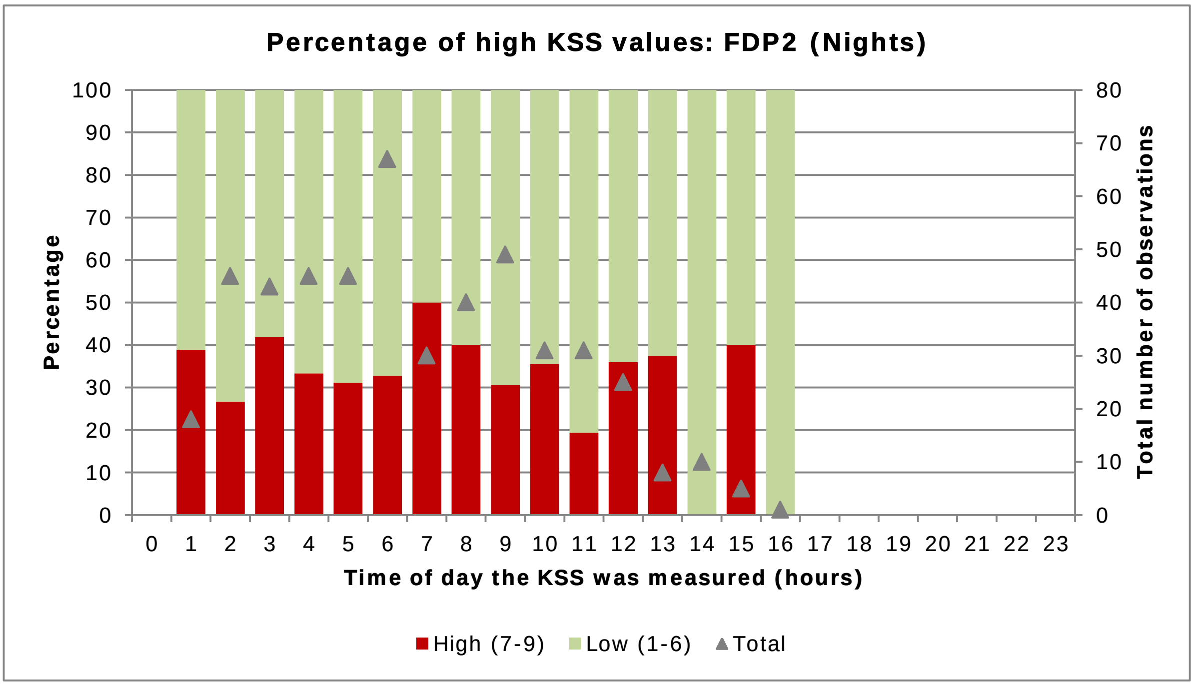


Figure 8 Percentages of high and low KSS TOD scores for each hour of the day for FDP2 (Nights)

Secondary objective for FDP2 (Nights)

The objective was to assess whether or not high fatigue scores during nights occur more frequently than in the FDP Baseline set.

The ratio estimates for the FDP2 (Nights) and the FDP Baseline set for the high KSS and SP scores were 1.394 and 1.565. Both estimates were larger than 1.0. This showed that high fatigue scores did occur more frequently in FDP2 (Nights) than in the FDP Baseline dataset.

Primary objective for consecutive FDP2 (Early starts)

The objective was to assess the prevalence of high levels of fatigue during consecutive (i.e., at least two in a row) early starts.

Table 26 presents occurrence-probability point estimates as well as 95% lower and upper confidence limits for KSS and SP. As can be seen from these estimates, high fatigue scores did occur for KSS TOD but not for SP TOD in the consecutive early starts as postulated under the hypotheses H_1 and H_0 respectively.

Table 26 KSS TOD and SP TOD occurrence-probability point estimates for consecutive FDP2 (Early starts)

Fatigue measure	N	*p*	*p*L	*p*U
KSS TOD 7, 8, 9	3	0.070	0.024	0.186
Nvalid	43			
SP TOD 6, 7	0	0.000	0.000	0.082
Nvalid	43			

Occurrence-probability point estimates (*p*) as well as 95% lower (*p*L) and upper confidence limits (*p*U).

The table below presents the descriptive statistics of the total sleep in 24 hours prior to TOD (Sleep24h) and Napping during FDP (FDPsleep) for the high (KSS ≥ 7) and low levels of fatigue (KSS < 7). The probability of sleep *not* to occur in 24 hours prior TOD is 0.312 for consecutive FDP2 (Early starts). The probability of a nap to occur during a duty is 0.013 for consecutive FDP2 (Early starts).

Table 27 Sleep24h and FDPsleep for high and low KSS TOD scores for consecutive FDP2 (Early starts)

		Min	***Max***	***Mean***	***SD***
Sleep24h	KSS High	250	440	322.78	73.829
(in min)	KSS Low	100	480	336.81	85.191
FDPsleep	KSS High	-	-	-	-
(in min)	KSS Low	25	25	25.00	-

SD: Standard Deviation.

The figure below presents the percentages of high and low fatigue scores (KSS) for each hour of the day.

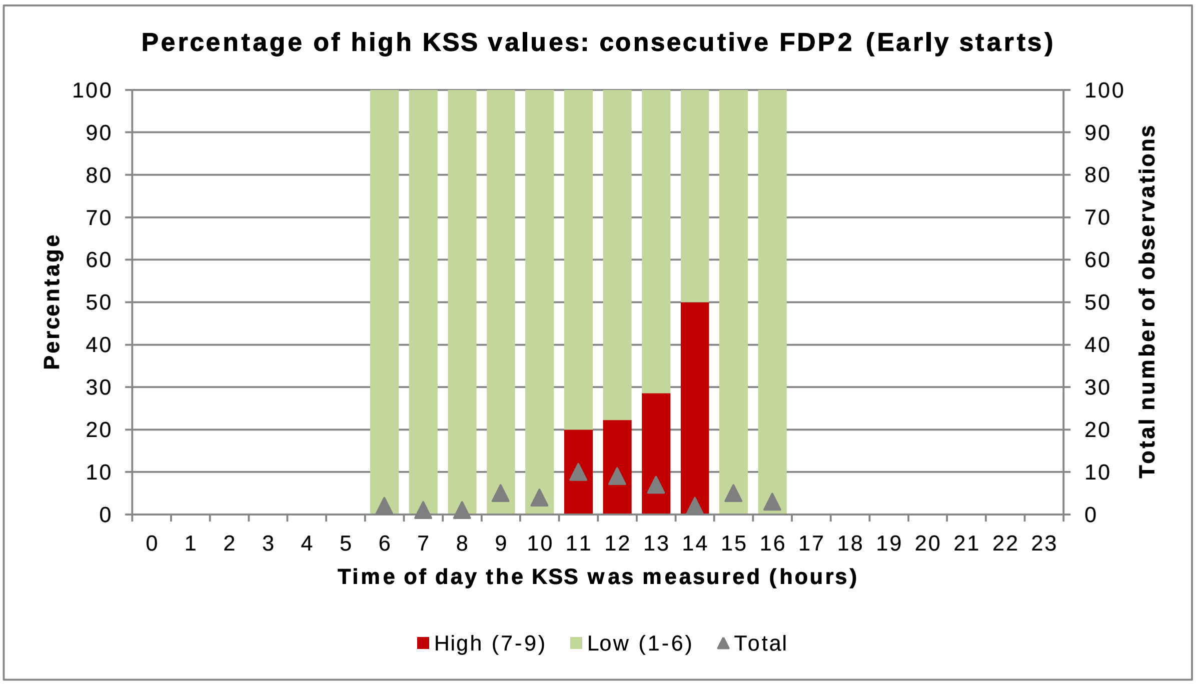


Figure 9 Percentages of high and low KSS TOD scores for each hour of the day for consecutive FDP2 (Early starts)

<u>Secondary objective for consecutive FDP2 (Early starts)</u>
The objective was to assess whether or not high fatigue scores during consecutive FDP2 (Early starts) occur more frequently than in the FDP Baseline set.

The ratio estimates for the consecutive FDP2 (Early starts) and the FDP Baseline set for the high KSS and SP scores were 0.290 and 0.000. Both estimates were smaller than 1.0. This showed that high fatigue scores did not occur more frequently in consecutive FDP2 (Early starts) than in the FDP Baseline dataset.

<u>Primary objective for consecutive FDP2 (Late finishes)</u>
The objective was to assess the prevalence of high levels of fatigue during consecutive (i.e., at least two in a row) late finishes.

Table 28 presents occurrence-probability point estimates as well as 95% lower and upper confidence limits for KSS and SP. As can be seen from these estimates, high fatigue scores did occur in the consecutive late finishes as postulated under the hypothesis H_1.

Table 28 KSS TOD and SP TOD occurrence-probability point estimates for consecutive FDP2 (Late finishes)

Fatigue measure	N	*p*	*p*L	*p*U
KSS TOD 7, 8, 9	5	0.294	0.133	0.531
Nvalid	17			
SP TOD 6, 7	3	0.176	0.062	0.410
Nvalid	17			

Occurrence-probability point estimates (*p*) as well as 95% lower (*p*L) and upper confidence limits (*p*U).

The table below presents the descriptive statistics of the total sleep in 24 hours prior to TOD (Sleep24h) and Napping during FDP (FDPsleep) for the high (KSS ≥ 7) and low levels of fatigue (KSS < 7). The probability of sleep *not* to occur in 24 hours prior TOD is 0.304 for consecutive FDP2 (Late finishes). The probability of a nap to occur during a duty is 0.000 for consecutive FDP2 (Late finishes).

Table 29 Sleep24h and FDPsleep for high and low KSS TOD scores for consecutive FDP2 (Late finishes)

		Min	*Max*	*Mean*	*SD*
Sleep24h	KSS High	370	580	462.86	74.714
(in min)	KSS Low	420	655	490.17	86.500
FDPsleep	KSS High	-	-	-	-
(in min)	KSS Low	-	-	-	-

SD: Standard Deviation.

The figure below presents the percentages of high and low fatigue scores (KSS) for each hour of the day.

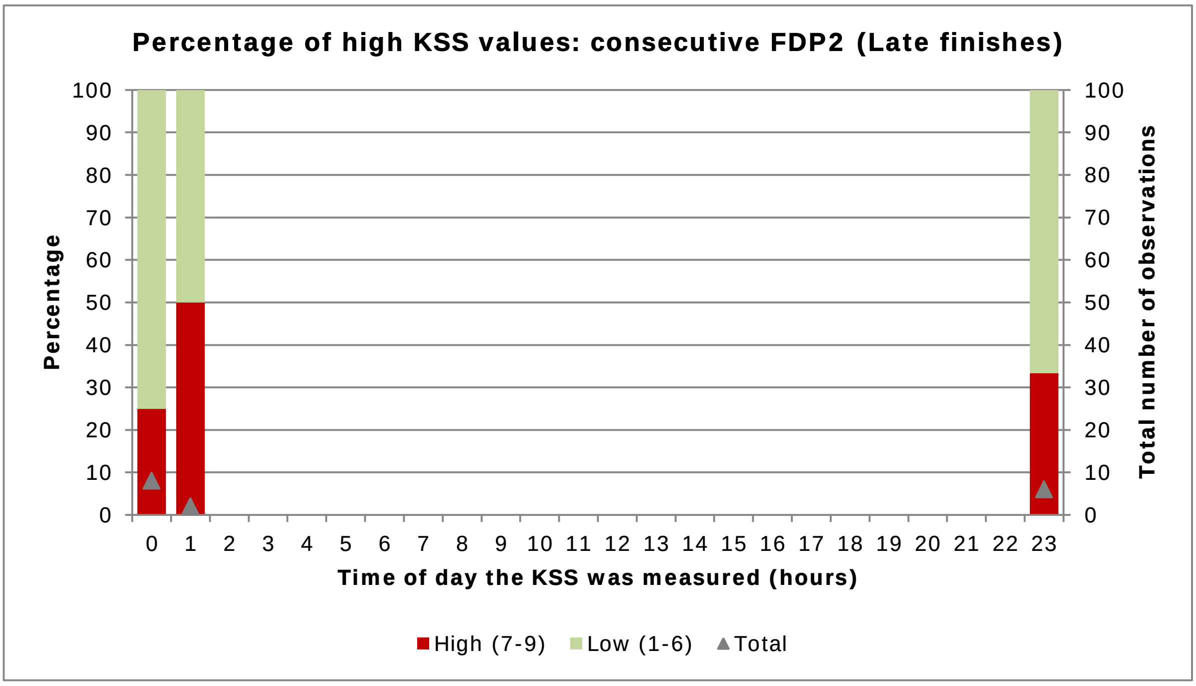


Figure 10 Percentages of high and low KSS TOD scores for each hour of the day for consecutive FDP2 (Late finishes)

<u>Secondary objective for consecutive FDP2 (Late finishes)</u>
The objective was to assess whether or not high fatigue scores during consecutive late finishes occur more frequently than in the FDP Baseline set.

The ratio estimates for the FDP Baseline set and the consecutive FDP2 (Late finishes) for the high KSS and SP scores were 1.220 and 1.914. Both estimates were larger than 1.0, showing that consecutive FDP2 (Late finishes) were more prone to high fatigue scores than the FDP Baseline set.

<u>Primary objective for consecutive FDP2 (Nights)</u>
The objective was to assess the prevalence of high levels of fatigue during consecutive (i.e., at least two in a row) nights.

Table 30 presents occurrence-probability point estimates as well as 95% lower and upper confidence limits for KSS and SP. As can be seen from these estimates, high fatigue scores did occur in the consecutive nights as postulated under the hypothesis H_1.

Table 30 KSS TOD and SP TOD occurrence-probability point estimates for consecutive FDP2 (Nights)

Fatigue measure	N	*p*	*p*L	*p*U
KSS TOD 7, 8, 9	32	0.348	0.258	0.449
Nvalid	92			
SP TOD 6, 7	14	0.151	0.092	0.237
Nvalid	93			

Occurrence-probability point estimates (*p*) as well as 95% lower (*p*L) and upper confidence limits (*p*U).

The table below presents the descriptive statistics of the total sleep in 24 hours prior to TOD (Sleep24h) and Napping during FDP (FDPsleep) for the high (KSS ≥ 7) and low levels of fatigue (KSS < 7). The probability of sleep *not* to occur in 24 hours prior TOD is 0.317 for consecutive FDP2 (Nights). The probability of a nap to occur during a duty is 0.065 for consecutive FDP2 (Nights).

Table 31 Sleep24h and FDPsleep for high and low KSS TOD scores for consecutive FDP2 (Nights)

		Min	***Max***	***Mean***	***SD***
Sleep24h	KSS High	30	455	284.12	109.989
(in min)	KSS Low	90	599	357.05	128.839
FDPsleep	KSS High	13	13	13.00	-
(in min)	KSS Low	25	35	30.00	4.082

SD: Standard Deviation.

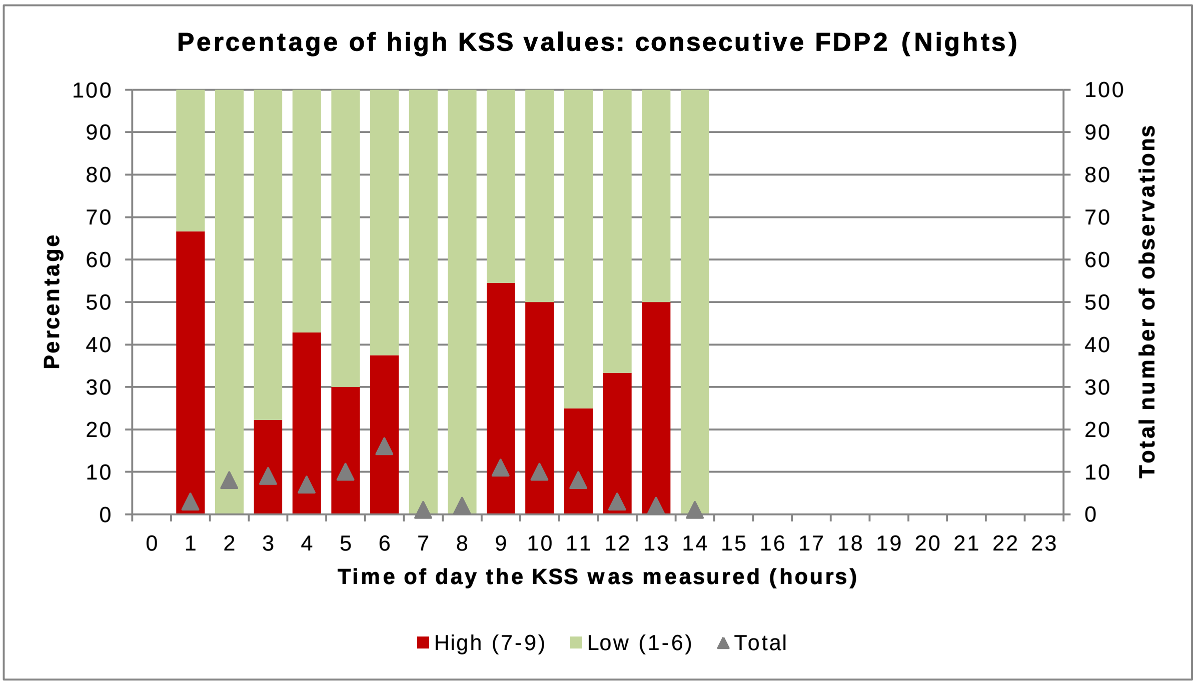


Figure 11 Percentages of high and low KSS TOD scores for each hour of the day for consecutive FDP2 (Nights)

The figure above presents the percentages of high and low fatigue scores (KSS) for each hour of the day.

Secondary objective for consecutive FDP2 (Nights)

The objective was to assess whether or not high fatigue scores during consecutive nights occur more frequently than in the FDP Baseline set.

The ratio estimates for the consecutive FDP2 (Nights) and the FDP Baseline set for the high KSS and SP scores were 1.443 and 1.633. Both estimates were larger than 1.0, showing that consecutive FDP2 (Nights) were more prone to high fatigue scores than the FDP Baseline set.

Primary objective for FDP2 (Mix)

The objective was to assess the prevalence of high levels of fatigue during a mix of early starts, late finishes and nights.

Table 32 presents occurrence-probability point estimates as well as 95% lower and upper confidence limits for KSS and SP. As can be seen from these estimates, high fatigue scores did occur in the mix of disruptive schedules as postulated under the hypothesis H_1.

Table 32 KSS TOD and SP TOD occurrence-probability point estimates for FDP2 (Mix)

Fatigue measure	N	*p*	*p*L	*p*U
KSS TOD 7, 8, 9	17	0.246	0.160	0.360
Nvalid	69			
SP TOD 6, 7	8	0.116	0.060	0.212
Nvalid	69			

Occurrence-probability point estimates (*p*) as well as 95% lower (*p*L) and upper confidence limits (*p*U).

The table below presents the descriptive statistics of the total sleep in 24 hours prior to TOD (Sleep24h) and Napping during FDP (FDPsleep) for the high (KSS ≥ 7) and low levels of fatigue (KSS < 7). The probability of sleep *not* to occur in 24 hours prior TOD is 0.062 for FDP2 (Mix). The probability of a nap to occur during a duty is 0.062 for FDP2 (Mix).

Table 33 Sleep24h and FDPsleep for high and low KSS TOD scores for FDP2 (Mix)

		Min	***Max***	***Mean***	***SD***
Sleep24h	KSS High	245	510	377.11	83.405
(in min)	KSS Low	170	515	378.21	95.470
FDPsleep	KSS High	17	17	17.00	-
(in min)	KSS Low	13	25	20.20	5.762

SD: Standard Deviation.

The figure below presents the percentages of high and low fatigue scores (KSS) for each hour of the day.

Secondary objective for FDP2 (Mix)

The objective was to assess whether or not high fatigue scores during a mix of early starts, late finishes and nights occur more frequently than in the FDP Baseline set.

The ratio estimates for FDP2 (Mix) and the FDP Baseline set for the high KSS and SP scores were 1.021 and 1.257. Both estimates were larger than 1.0, showing that consecutive FDP2 (Mix) were more prone to high fatigue scores than the FDP Baseline set.

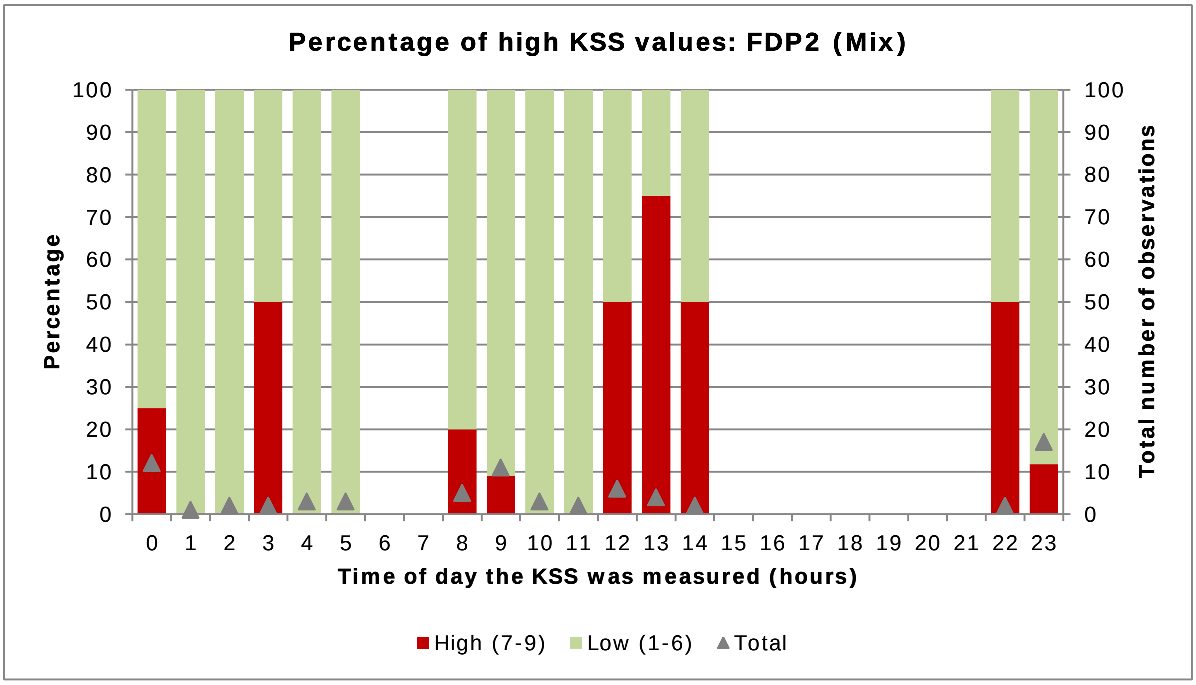


Figure 12 Percentages of high and low KSS TOD scores for each hour of the day for FDP2 (Mix)

INTERIM CONCLUSION

The goal of this step was to assess the prevalence of high fatigue scores during the different types of (consecutive) disruptive flight duties.

What we have learned from this step is that high fatigue scores do occur for most types of disruptive schedules, consecutive or not.

The proportion of high fatigue is higher than in the FDP Baseline set for most non-consecutive and consecutive disruptive schedules, with early starts and consecutive early starts as exceptions to this. Note that the same FDP Baseline set – with the same implications – was used for these disruptive FDPs as for the FDP1 (Night duties of more than 10 hours).

The results for consecutive early starts and consecutive late finishes should be interpreted with caution given the relatively small sample size for this particular sub-dataset.

Compare fatigue scores between FDP categories

The main approach of the following analyses was to compare a certain type of disruptive FDP with non-disruptive (all FDPs with start time ≥ 07:00h and end time < 23:00h – also referred to as daytime) FDPs. Disruptive types consist of early starts, late finishes, and nights.

The method of analysis was exactly the same as for the analysis of long- vs short-night FDPs, except for the analysis of consecutive FDPs of the same type. For these FDPs the first FDP of, for example an early start, was compared with the immediately following FDP of the same type.

The variables used were similar to the ones used to compare fatigue scores between long- and short-night FDPs. In addition, we computed the mean and Anova F-ratios for the two FDP categories for all the other variables. This was done in order to illustrate in what ways the two FDP categories differed on the other predictors of high levels of fatigue.

Comparison between early start and non-disruptive FDPs

Figure **13** shows the mean KSS ratings at TOD for the early start (05:00h - 06:59h) and non-disruptive FDPs in the between-subjects data. Descriptive statistics showed the KSS mean was higher for the early start than for the non-disruptive FDPs, with the mean difference being 0.4 units. The mean KSS ratings were very close to 5 for the early start FDPs and to 4.5 for the non-disruptive one. Note that no statistical tests or adjustments for other factors were applied.

Table 34 shows a significant OR for early start/daytime FDPs vs high levels of fatigue. The value indicates that an early start FDP is 2 times more likely to show high levels of fatigue than a daytime FDP. For Function, the OR was not significant. All other predictors were added one at a time separately, but none affected the OR appreciably.

Table 34 Results from logistic regression predicting high levels of fatigue from early start/daytime FDPs and Function. N = 294

Predictors	***OR***	***95% CI***	***p***	***% high fatigue***
Early start/daytime	2.019	1.088; 3.743	.026	23/14
Pilot/cabin	1.396	0.662; 2.945	.381	

OR = Odds Ratio. CI = 95% Confidence Interval. p = level of significance.
The result is adjusted for Age and Gender. % high fatigue = % with KSS ≥ 7 in the early start/daytime FDPs.

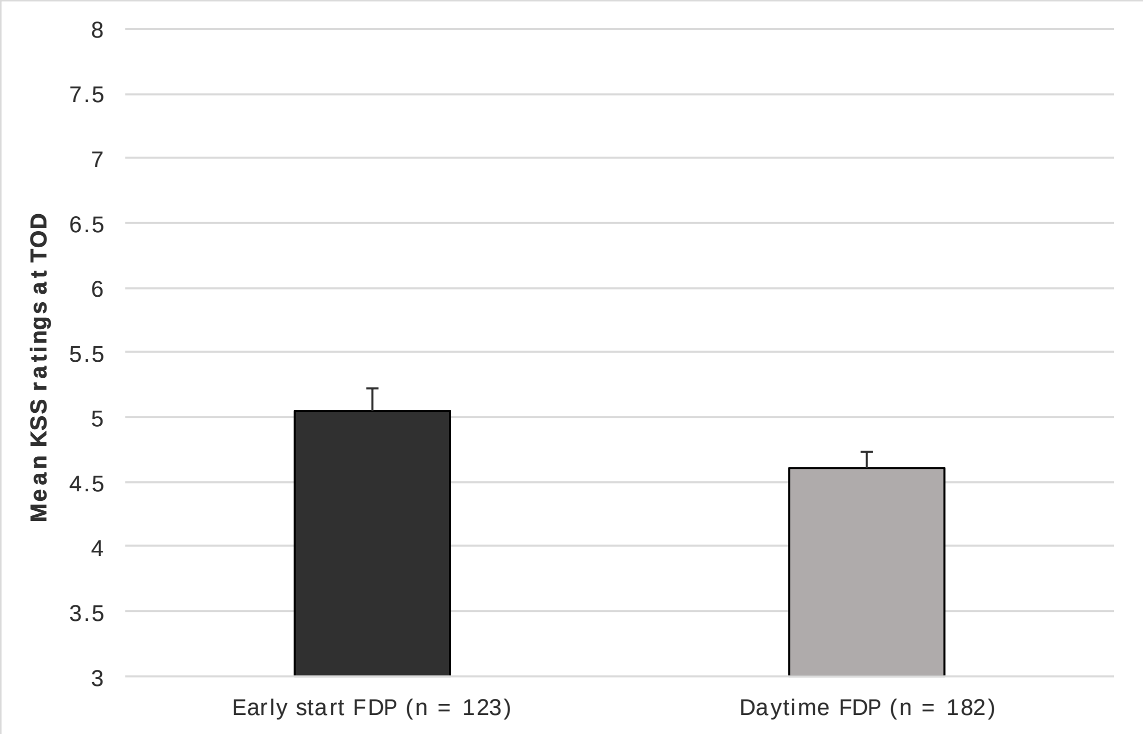


Figure 13 Mean KSS ratings at TOD for the early start (black bars) and daytime (non-disruptive; grey bars) FDPs in the between-subjects data. The vertical lines denote the standard errors. Note that the y-axis covers only part of the 9-point scale

Table 35 shows that the means of the early start and daytime FDPs differed, as expected, with respect to the timing variables. The early group also showed more sectors flown, a shorter time awake, less sleep, less napping, and a higher number of time zones flown. FDP duration did not differ.

Table 35 Anova, mean±SD for daytime (non-disruptive) and early start FDPs. No adjustments. N = 305

	Daytime *Mean±SD*	***Early start*** *Mean±SD*	***F-ratio***	***p***
FDPWOCLm: Time in WOCL			-	
FDPsectors: Sectors flown	1.84±0.95	2.54±1.06	35.59	.000
AwakeTOD: Time awake prior TOD	13.11±5.72	11.58±4.32	6.23	.013
EndH: FDP end time	19.14±3.93	12.77±2.40	258.24	.001
FDPdur: FDP duration	7.42±3.14	7.15±2.28	0.66	.417
Sleep24h: Sleep in 24h prior TOD	7.03±2.13	5.44±1.85	44.86	.000
TimeDayTOD: Time of day at TOD	18.18±3.94	11.85±2.41	252.98	.000
StartH: FDP start time	11.86±3.93	11.85±3.95	304.43	.000
FDPsleep: Napping during FDP	2.60±7.47	1.67±5.59	4.03	.046
TZ_EW: Time zones crossed EW	1.30±2.60	0.31±1.01	147.37	.000
TZ_WE: Time zones crossed WE	1.12±2.18	0.47±1.48	14.48	.000

SD = Standard Deviation. p = level of significance.

When using the maximum value of KSS instead of that obtained at TOD of the final leg, we found for early start/daytime OR = 1.78 (CI = 1.02; 3.08) and p = .041. For Function we found OR = 1.19 (CI = 0.67; 2.11) and p = .561. N = 304. The values are similar to that in Table 34.

INTERIM CONCLUSION

The goal of this step was to compare the occurrence of high fatigue at TOD between the early start and daytime (non-disruptive) FDPs using between-subjects data.

Our interim conclusion is that the odds of high fatigue are significantly increased at TOD during the early start FDPs as compared to the daytime FDPs.

Comparison between late finish and non-disruptive FDPs

Figure 14 shows the mean KSS ratings at TOD for the late finish (23:00h - 01:59h) and non-disruptive FDPs in the between-subjects data. The KSS mean was higher for the late finish than for the non-disruptive FDPs, with the mean difference being 0.9 units. The mean KSS ratings fell between 5.0 and 5.5 for the late finish FDPs. For the non-disruptive FDPs, the mean ratings were close to 4.5 and 5.0. Note that no statistical tests or adjustments for other factors were applied.

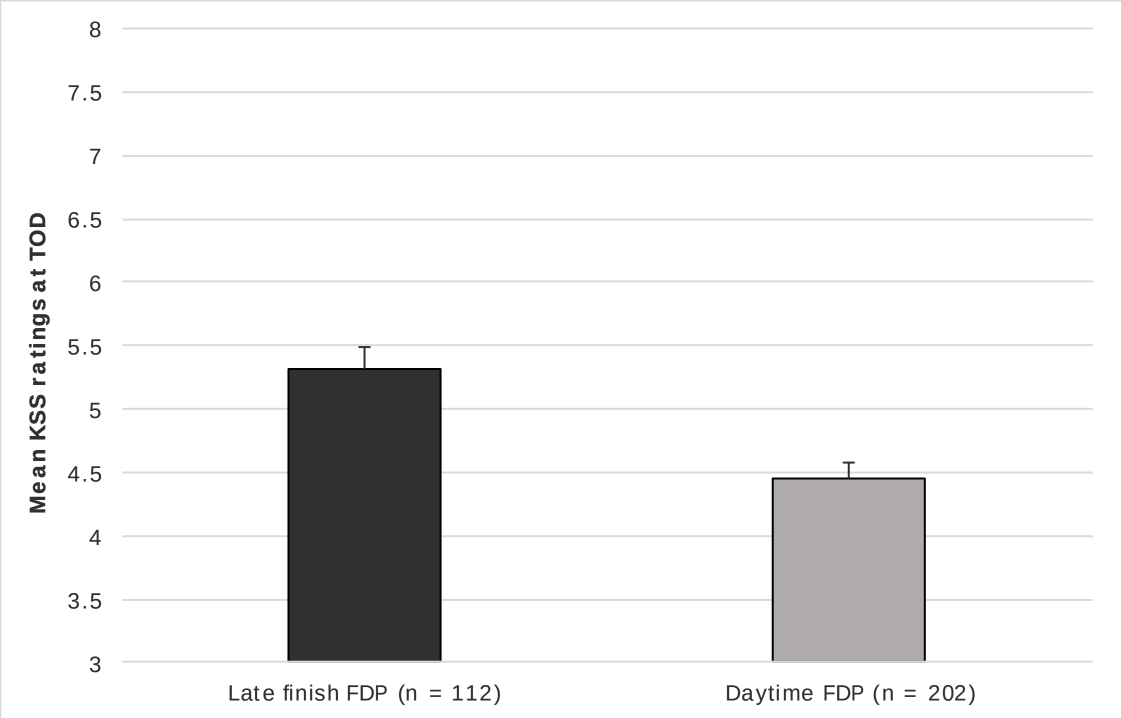


Figure 14 Mean KSS ratings at TOD for the late finish (black bars) and daytime (non-disruptive; grey bars) FDPs in the between-subjects data. The vertical lines denote the standard errors. Note that the y-axis covers only part of the 9-point scale

Table 36 shows a significant OR for late finish/daytime high levels of fatigue. The value indicates that a late FDP is 3.8 times more likely to show high levels of fatigue than a daytime FDP; i.e., a strongly significant effect. For Function, the OR was not significant. All other predictors were added one at a time separately, but none affected the OR appreciably. The percentage of crew members with high levels of fatigue was almost three times higher in the late finish FDPs.

Table 36 Results from multiple logistic regression analysis predicting high levels of fatigue from late finish/daytime FDPs and Function. N = 294

Predictors	*OR*	*95% CI*	*p*	*% high fatigue*
Late finish/daytime	3.772	2.066; 6.888	.000	32/13
Pilot/cabin	1.776	0.868; 3.631	.116	

OR = Odds Ratio. CI = 95% Confidence Interval. p = level of significance.
The result is adjusted for Age and Gender. % high fatigue = % with KSS ≥ 7 in the late finish/daytime FDPs.

Table 37 shows that the timing variables differed, as expected, between the FDPs. Among the other variables the late finish FDPs show fewer sectors, more time awake, longer FDP duration, shorter sleep, and more time zones crossed West to East.

Table 37 Anova, mean±SD for night and daytime (non-disruptive) FDPs. No adjustment. N = 314

	Daytime *Mean±SD*	***Night*** *Mean±SD*	***F-ratio***	***p***
FDPWOCLm: Time in WOCL	-	-		
FDPsectors: Sectors flown	2.10±1.04	1.79±0.85	6.90	.009
AwakeTOD: Time awake prior TOD	12.82±2.15	16.65±6.91	30.88	.000
EndH: FDP end time	18.46±3.91	00.46±0.73	257.52	.000
FDPdur: FDP duration	7.26±2.95	8.35±3.04	9.66	.002
Sleep24h: Sleep in 24h prior TOD	7.25±2.12	6.50±2.42	7.98	.005
TimeDayTOD: Time of day at TOD	17.50±3.91	23.46±0.48	253.61	.001
StartH: FDP start time	11.20±3.72	16.11±3.18	138.71	.000
FDPsleep: Napping during FDP	2.13±6.19	2.55±6.90	0.27	.606
TZ_EW: Time zones crossed EW	0.97±2.29	0.89±2.21	0.09	.771
TZ_WE: Time zones crossed WE	0.84±1.89	1.20±2.38	12.07	.001

SD = Standard Deviation. p = level of significance.

When maximum KSS was used as outcome instead of KSS at TOD for the analysis in Table 36, we found for late finish OR = 2.98 (CI = 1.74; 5.10) with p = .000. For Function we obtained OR = 1.18 (CI = 0.67; 2.06) with p = .570. N = 312. The OR was somewhat lower than that obtained with KSS at TOD as outcome.

INTERIM CONCLUSION

The goal of this step was to compare the occurrence of high fatigue at TOD between the late finish and daytime (non-disruptive) FDPs using between-subjects data.

Our interim conclusion is that the odds of high fatigue are significantly increased at TOD during the late finish FDPs as compared to the daytime FDPs.

Comparison between night and non-disruptive FDPs

Figure 15 shows descriptive statistics for the night (FDP encroaching on the WOCL) and non-disruptive FDPs in the between-subjects data. The KSS mean was higher for the night than for the non-disruptive FDPs, with the mean difference being 0.8 units. The mean KSS ratings were close to 5.5 for the night FDPs and 4.5 for the non-disruptive one. Note that no statistical tests or adjustments for other factors were applied.

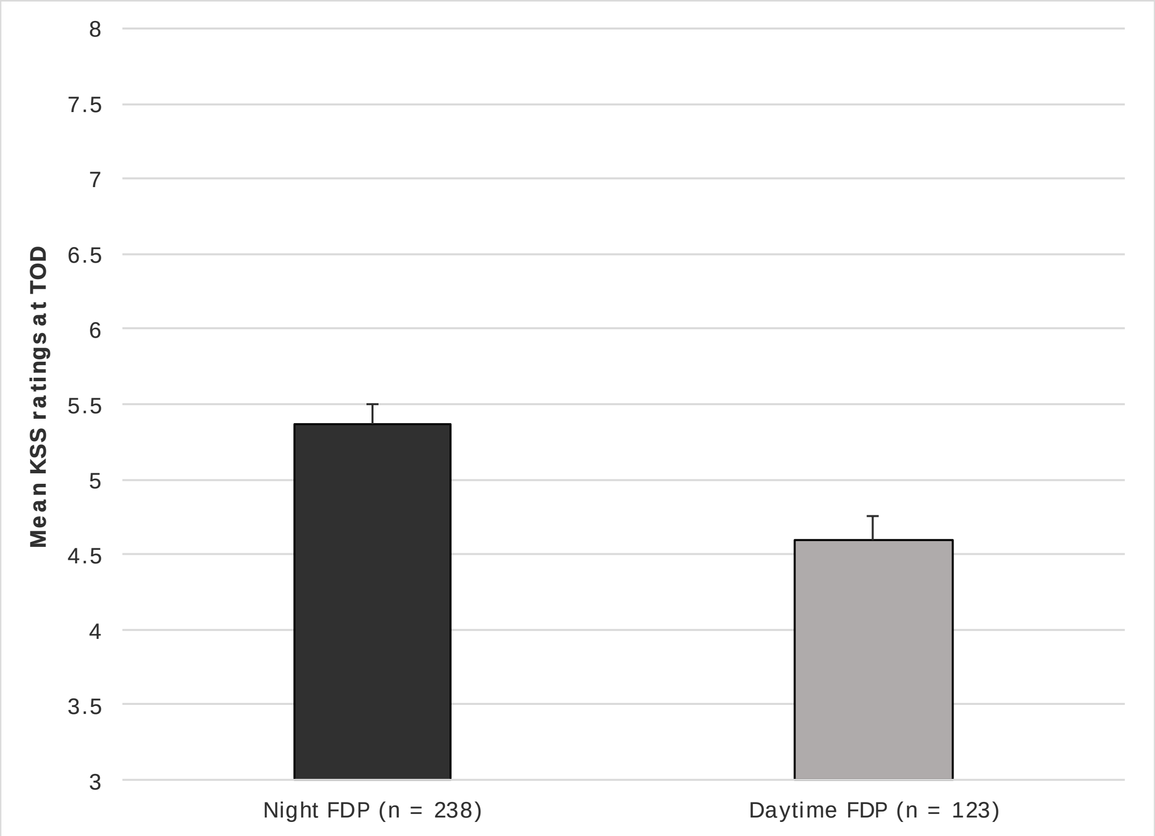


Figure 15 Mean KSS ratings at TOD for the night (black bars) and daytime (non-disruptive; grey bars) FDPs in the between-subjects data. The vertical lines denote the standard errors. Note that the y-axis covered only part of the 9-point scale

Table 38 shows a significant OR for night/daytime vs high levels of fatigue. The value indicates that a night FDP is 3.2 times more likely to show high levels of fatigue than a daytime FDP. For Function, the significant OR indicates that cabin crew is 1.8 times more likely to show high levels of fatigue than pilots. All other variables were added one at a time separately, but none affected the OR appreciably. The percentage with high levels of fatigue was highest in the night FDP group.

Table 38 Results from logistic regression predicting high levels of fatigue from night/daytime FDPs and Function. N = 348

Predictors	***OR***	***95% CI***	***p***	***% high fatigue***
Night/daytime	3.206	1.805; 5.695	.000	34/15
Pilot/cabin	1.763	1.072; 2.899	.026	

OR = Odds Ratio. CI = 95% Confidence Interval. p = level of significance.
The result is adjusted for Age and Gender. % high fatigue = % with KSS ≥ 7 in the night/daytime FDPs.

Table 39 shows the means for the other variables of the night and daytime FDPs. The variables related to timing of the FDPs are, as expected, extremely different between the night and daytime FDPs. Among other differences, it is worth noting that night FDPs involve fewer sectors, more time zones crossed, a longer time awake, a longer FDP duration, less sleep, and more napping.

Table 39 Anova, mean±SD for night and daytime (non-disruptive) FDPs. No adjustment. N = 361

	Daytime *Mean±SD*	***Night*** *Mean±SD*	***F-ratio***	***p***
FDPWOCLm: Time in WOCL	0	2.68±1.24	-	
FDPsectors: Sectors flown	2.51±1.07	1.71±0.947	52.65	.000
AwakeTOD: Time awake prior TOD	12.91±4.95	15.38±7.44	10.86	.000
EndH: FDP end time	18.59±4.10	7.56±3.65	678.41	.000
FDPdur: FDP duration	7.21±2.03	8.43±8.02	11.93	.000
Sleep24h: Sleep in 24h prior TOD	7.14±2.03	4.48±2.68	82.25	.000
TimeDayTOD: Time of day at TOD	17.66±4.10	6.63±3.67	675.04	.000
StartH: FDP start time	11.37±3.64	23.13±3.89	773.55	.000
FDPsleep: Napping during FDP	1.89±6.91	4.14±9.07	5.85	.016
TZ_EW: Time zones crossed EW	0.46±1.55	1.04±2.11	6.65	.011
TZ_WE: Time zones crossed WE	0.47±1.50	2.09±2.98	28.94	.000

SD = Standard Deviation. p = level of significance.

When maximum KSS was used instead of KSS at TOD in the analysis of night FDPs vs daytime controls in Table 40, we found OR = 3.28 (1.90; 5.53) with p = .000. For Function we obtained OR = 1.56 (0.96; 2.52) with p = .072. N = 360. The results are very similar to those using KSS at TOD as outcome.

INTERIM CONCLUSION

The goal of this step was to compare night and daytime (non-disruptive) FDPs with respect to the occurrence of high fatigue at TOD using between-subjects data.

Our interim conclusion is that the odds of high fatigue are significantly increased at TOD during the night FDPs as compared to the daytime FDPs.

Comparison between first and second (consecutive) disruptive FDPs

First and second (consecutive) disruptive FDPs were examined through the pertinent within-subject data. The within-subjects analyses used the first FDPs that fulfilled the criterion for *both* FDP types *within the same participant*.

Figure 16 shows descriptive statistics for the first and second (consecutive) early starts, late finishes, and nights based on this data. The mean KSS ratings fell between 4.5 and 5.5 in both the first and second FDP categories and no increase in level of fatigue was observed during the second FDP in a row. Note that no statistical tests or adjustments for other factors were applied.

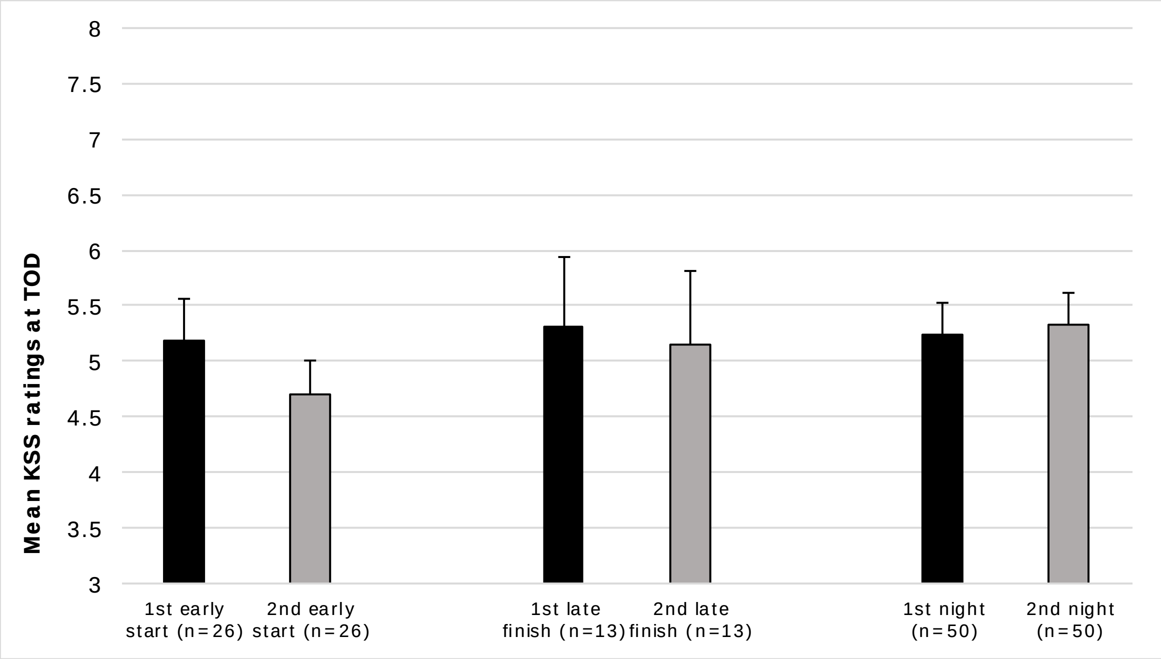


Figure 16 Mean KSS ratings at TOD for the first (black bars) and second (consecutive) (grey bars) night FDPs in the within-subject data. The vertical lines denote the standard errors. Note that the y-axis covers only part of the 9-point scale

In order to analyse the effects of consecutive disruptive FDPs we applied a conditional (fixed effects) logistic regression comparing the second FDP with the first FDP (see Table 40). We did not analyse late finish FDPs due to low a low N. None of the other analyses (on early starts and nights) showed any significant difference. The introduction of Function did not affect the results; neither did the introduction of any of the covariates.

We interpret the results as a lack of evidence of a difference between a first and a second consecutive FDP of the same type, but with reservations due to few observations. No generalization can be made to longer sequences. The latter were extremely rare.

It might be of interest to compare the amount of prior sleep between the two consecutive early starts. It was 5.64±1.44h (mean±SD) and 6.44±1.31h, respectively ($p < .05$). One possible effect is that the longer sleep before the second early start may have counteracted fatigue.

Table 40 Comparison of the first and immediately following disruptive FDP for early, starts, late finishes, and night flights. Conditional logistic regression, with covariates added one at a time

Main variables	***OR***	***95% CI***	***p***	***N***
Consecutive early starts 1/2	0.50	0.09; 2.73	.423	26
Consecutive late finishes 1/2	-	-	-	13
Consecutive nights 1/2	1.30	0.57; 2.96	.533	50

OR = Odds Ratio. CI = 95% Confidence Interval. p = level of significance.
All variables entered (including Function, Gender and Age), but not displayed if not significant.

INTERIM CONCLUSION

The goal of this step was to compare the first and consecutive FDPs of the same type with respect to fatigue at TOD using within-subjects data. These schedules consisted of two consecutive early start, late finish or night FDPs.

We concluded that fatigue does not seem to accumulate over two consecutive early starts and two consecutive nights. Accumulation of fatigue over two consecutive late finish FDPs remains open due to limited data.

Find clusters of variables: Prediction of high fatigue

To contribute to the understanding of fatigue hotspots, we carried out a series of logistic regression analyses to predict the OR for high levels of fatigue for each type of FDP separately; i.e., for early starts, late finishes and nights. First, all the simple logistic regressions are presented in model 1. In model 2, only the significant predictors from model 1 are included, except for the disruptive/non-disruptive predictor. In case of collinearity additional models are presented excluding specific variables.

Predicting fatigue from all predictors: early start / non-disruptive dataset

All predictors were entered in model 1. Only FDP start time was a significant predictor in model 1 (Table 41). High fatigue becomes more prevalent with earlier start. Note that only start times between 05:00h and 06:59h are included in early start FDPs. We interpret the results as indicating a marginal effect of early starts on high fatigue.

Table 41 Results from logistic regression of predictors vs a high level of fatigue. Early start/non-disruptive dataset. Model 1 = simple logistic regression – each predictor analysed separately. No multiple regression analysis was carried out since only one predictor was significant in the model 1. N = 299.

	Model 1 *Unadjusted OR (CI)*	***p***
FDPdur: FDP duration	1.11 (0.99;1.23)	.065
FDPWOCL	#	
FDPsectors: No. of sectors	1.04 (0.78;1.37)	.803
AwakeTOD: Time awake prior TOD	1.05 (1.00;1.11)	.051
EndH: FDP end time	0.98 (0.92;1.04)	.529
Sleep24h: Sleep in 24h prior TOD	0.90 (0.78;1.02)	.103
StartH: FDP start time	0.92 (0.85;0.99)	.045
Function (1 pilot/2 cabin)	1.36 (0.74;2.50)	.321
Gender (1 male/2 female)	0.96 (0.49;1.88)	.957
Age (in years)	0.98 (0.95;1.01)	.212
FDPsleep: Napping during FDP	0.99 (0.95;1.04)	.788
TZ_WE: Time zones crossed from West to East	1.17 (1.00;1.34)	.051
TZ_EW: Time zones crossed from East to West	0.94 (0.80;1.11)	.474
FDP2ES: FDP2 early start	1.77 (0.98;3.20)	.059

OR = Odds Ratio. CI = 95% Confidence Interval. # = removed variable. p = level of significance.

<u>Excluding the non-disruptive FDP controls or early start FDPs</u>

We made a separate analysis of early start FDPs alone. None of the predictors showed significant results.

INTERIM CONCLUSION

The somewhat higher odds of high fatigue at TOD during early start FDPs, as compared to non-disruptive (daytime) FDPs, were marginally due to an earlier start time itself.

When early start FDPs were considered alone, none of the FDP-related characteristics studied (including prior sleep) show significant results.

Predicting fatigue from all predictors: late finish / non-disruptive dataset

All predictors were entered in model 1, except for FDPWOCL (by definition no encroachment on the WOCL can occur during late finishing FDPs). Late finish FDPs contain those that finish between 23:00h and 01:59h). Table 42 shows strong effects for several predictors in model 1, including the late finish/daytime predictor per se, FDP duration, Time awake prior to TOD, Sleep in 24 hours prior to TOD, and Time zones crossed from West to East. In model 2 only significant predictors were included, except for late finish/daytime FDP type. The results show (long) FDP duration was associated with a higher likelihood of high fatigue. Note that end times are very different between late finish FDPs and daytime controls. It is, therefore, not a suitable variable for inclusion in the analysis when both groups are included. Also note that the regressions include influence from daytime controls.

Table 42 Results from logistic regression of predictors vs a high level of fatigue. Late finish/non-disruptive dataset. Model 1 = simple logistic regression – each predictor analysed separately. Models 2 = multiple predictors entered together. N = 314

	Model 1 *Unadjusted OR (CI)*	***p***	***Model 2*** *OR (CI) Non-sign and late finish/ daytime not entered*	***p***
FDPdur: FDP duration	1.23 (1.11;1.37)	.000	1.13 (1.01;1.27)	.038
FDPWOCL	#		#	
FDPsectors: No. of sectors	1.17 (0.89;1.54)	.261		
AwakeTOD: Time awake prior TOD	1.09 (1.04;1.14)	.000	1.02 (0.97;1.06)	.499
Sleep24h: Sleep in 24h prior TOD	1.17 (1.08;1.28)	.000		
StartH: FDP start time	1.03 (0.96;1.10)	.399		
Function (1 pilot/2 cabin)	1.45 (0.82;2.58)	.201		
Gender (1 male/2 female)	0.90 (0.48;1.69)	.744		
Age (in years)	0.90 (0.48;1.69)	.744		
FDPsleep: Napping during FDP	0.98 (0.94;1.03)	.385		
TZ_WE: Time zones crossed from West to East	1.16 (1.05;1.30)	.005	1.07 (0.95;1.21)	.237
TZ_EW: Time zones crossed from East to West	0.88 (0.74;1.03)	.116		
FDP2LF: FDP2 late finish	3.21 (1.81;5.68)	.000	#	

WOCL not entered since late finishes do not include the WOCL.
OR = Odds Ratio. CI = 95% Confidence Interval. # = removed variable. p = level of significance.

Excluding the non-disruptive controls

As with the early start FDP data, we undertook a separate analysis of late finish FDPs alone in order to exclude influence from daytime controls. All predictors from the main analysis were used in the simple logistic regression analysis, except for the late finish/non-disruptive variable (which was not meaningful in this analysis). Three variables became significant: FDP duration (OR = 1.20 (CI = 1.04; 1.39), p = .012), FDP start time (OR = 0.82 (CI = 0.71; 0.94), p = .004), FDP end time (OR = 0.54 (CI = 0.31; 0.95), p = .033). Entering all three together gives an unstable solution because of collinearity. The impression of the analyses is that early start time, early end time, and long FDP duration increase the odds of high fatigue. Notably, prior sleep duration was not a significant predictor, which seems logical since late finish FDPs should not interfere with prior sleep (N = 112). Note that end time does not constitute a statistical problem when the non-disruptive controls are excluded.

Since the OR for FDP end time was, unexpectedly, significantly less than 1, the result indicates that earlier end time involves higher fatigue. To check on this result we computed the proportions of high fatigue for the three times. This yielded for 23:01h - 24:00h = 0.57, for 24:01h - 01:00h = 0.23, and for 01:01h - 02:00h = 0.30 ($p = .015$). Thus, fatigue was considerably higher at 23:01h - 24:00h. We then calculated an analysis of variance between end times and all other predictors. This yielded a significant F-ratio for FDP duration ($F = 6.1$, $p = .003$,) with mean±SE for end times 23:01h - 24:00h = 9.7±.4h, 24:01h - 01:00h = 7.4±.4h, and 01:01h - 02:00h = 9.0±.4h. N = 23, 56, and 33 for the three times, respectively. The results suggest, but do not prove, that long FDP duration may have influenced the high fatigue results. It should be emphasized that we only have three measurement points for the analysis of effect of end times, and that this limits the possibility to draw conclusions.

<u>Excluding late finish FDPs</u>

Analysing the non-disruptive control group alone yielded significant simple ORs for Time awake prior TOD (OR = 1.09 (CI = 1.01; 1.17), $p = .022$), FDP duration (OR = 1.21 (CI = 1.03; 1.40), $p = .017$), and FDP end time (OR = 1.13 (CI = 1.01; 1.28), $p = .046$). When all are entered in the same multiple logistic regression analysis none of the predictors come out significant. Thus, even if long time awake, long FDP duration and late end time are significant predictors of high fatigue, their impact is quite weak and they are inter-correlated.

INTERIM CONCLUSION

The higher odds of high fatigue at TOD during late finish FDPs, as compared to non-disruptive (daytime) FDPs, were mainly due to a longer FDP duration, in addition to a late end time itself.

When late finish FDPs were considered alone, FDP duration was the most evident FDP-related characteristic to predict which of these FDPs involve high fatigue at TOD (the longer the duration, the higher the probability of high fatigue). Other predictors were early start time and early end time.

Predicting fatigue from all predictors: night/non-disruptive dataset

All predictors were entered in model 1. The night/daytime predictor, FDPWOCL and Sleep in 24 hours prior TOD were the strongest predictors, together with FDP start time. The significant variables from the simple logistic regression analysis were then entered in model 2, except for night/daytime (which represents the FDP type). In that model we found strong collinearity (high correlations) between FDPWOCL, FDP start time, and FDP end time, Collinearity causes inflation of variance as well unstable solutions. The main conflict was between FDPWOCL and StartH (correlation r = 0.85). We therefore computed model 2 leaving out StartH. This left FDPWOCL and Sleep24h. In model 3 we instead removed WOCL. This leads to Sleep in 24 hours prior TOD becoming the strongest predictor. In model 4 we wanted to show which predictors would take the place of sleep, when that predictor was removed (as well as WOCL). FDP duration and FDP start time then appeared instead. Note that end times are very different between late finish FDPs and non-disruptive controls. It is, therefore, not a suitable variable for inclusion in the analysis when both groups are included. Also note that the regressions include influence from non-disruptive controls.

Table 43 Results from logistic regression of predictors vs high fatigue at TOD. Night/non-disruptive dataset. Model 1 = each predictor analysed separately. Models 2-4 multiple predictors entered together. N = 361-335 depending on variable

	Model 1 *Unadjusted OR (CI)*	***p***	***Model 2*** *OR (CI) Non-sign and FDP2N not entered*	***p***	***Model 3*** *OR (CI) WOCL removed*	***p***	***Model 4*** *OR (CI) WOCL and sleep removed*	***p***
FDPdur: FDP duration	1.08 (1.01;1.16)	.037	0.97 (0.88;1.07)	.561	1.05 (0.96;1.14)	.295	1.10 (1.02;1.18)	.014
FDPWOCL	3.00 (1.69;5.33)	.000	3.20 (1.11;9.19)	.047	#		#	
FDPsectors: No. of sectors	0.91 (0.73;1.14)	.428						
AwakeTOD: Time awake prior TOD	1.03 (0.99;1.06)	.090						
Sleep24h: Sleep in 24h prior TOD	0.84 (0.76;0.91)	.000	0.86 (0.77;0.96)	.009	0.88 (0.79;0.99)	.016	#	
StartH: FDP start time	1.06 (1.02;1.10)	.003	#		1.04 (1.00;1.09)	.070	1.05 (1.01;1.09)	.022
Function (1 pilot/ 2 cabin)	1.59 (0.98;2.57)	.059						
Gender (1 male/ 2 female)	0.78 (0.47;1.32)	.356						
Age (in years)	1.00 (0.97;1.02)	.878						
FDPsleep: Napping	1.00 (0.98;1.03)	.877						
TZ_WE: Time zones crossed WE	1.04 (0.95;1.14)	.426						
TZ_EW: Time zones crossed EW	1.06 (0.94;1.20)	.336						
FDP2N: FDP2 night	3.03 (1.72;5.34)	.000	#		#		#	

OR = Odds Ratio. CI = 95% Confidence Interval. # = removed variable. p = level of significance.

The overall impression is that the key factor behind fatigue in night FDPs is the WOCL. A late start time makes an FDP end time in the WOCL more likely. The case is similar for FDP duration in the present context of late finish/night flights. In addition, FDPs encroaching on the WOCL had a proportion of 0.34 of FDPs with a high level of fatigue. For FDPs not encroaching on the WOCL, the proportion was 0.15 ($p = .000$).

Excluding the non-disruptive controls

The use of both night FDP and non-disruptive controls used above was necessary to obtain a contrast for evaluating the overall importance of the WOCL. However, it is also of interest to evaluate the prediction of high fatigue within the night FDP group alone. Thus, the same analysis as above was carried out using all the significant predictors from the simple logistic regression analysis (not presented), except for the night/daytime FDP variable itself and WOCL (not meaningful in this analysis since it is part of the definition of night FDPs). The significant variables became: Sleep in 24 hours prior TOD (OR = 0.84 (CI = 0.75; 0.94), $p = .002$) and Function (OR = 1.89 (CI = 1.07; 3.36), $p = .029$). When the two significant variables were entered into a multiple logistic regression analysis, Sleep in 24 hours prior TOD (OR = 0.79 (CI = 0.67; 0.93), $p = .004$) was significant, and Function (OR = 2.33 (CI = 0.99; 5.44), $p = .051$) almost significant. No other predictor showed a significant OR when sleep was removed. The same analysis was carried out with hours within the WOCL instead of the variable '≥ 1 minute in the WOCL' used in the other analyses. The same result was obtained – only Sleep in 24 hours and Function were significant predictors (N = 189).

Excluding night FDPs

To further understand the data we carried out a similar analysis for non-disruptive control FDPs only. The simple logistic regression analyses were significant for Number of sectors (OR = 1.67 (CI = 1.03; 2.71), $p = .039$), and Age (OR = 0.94 (0.88; 0.99), $p = .045$). Fatigue increased with sectors flown and with younger age. No variables were significant in the multiple regression analysis. Note that FDP duration is not significant in this analysis or in the analysis of the night only dataset.

INTERIM CONCLUSION

The higher odds of high fatigue at TOD during night FDPs, as compared to non-disruptive (daytime) FDPs, were mainly due to encroaching on the WOCL. Longer FDP duration and later FDP start time also played a role. Besides these FDP characteristics, shorter prior sleep and being a cabin crew member (vs a pilot) were associated with higher odds of high fatigue.

When night FDPs were considered alone only the amount of sleep in the past 24 hours significantly predicted which of these FDPs involve a higher likelihood of fatigue at TOD (WOCL was not included as a predictor since all included night FDPs include encroachment on the WOCL). Longer prior sleep was associated with lower likelihood of high fatigue.

Analyses of the variable late finish plus night FDPs

The analyses of early start, late finish, and night FDPs were based on the definitions in the current FTL regulations. While working with the data it became obvious that late finish and night FDPs represents a continuum of lateness and that late finish flights may end close to the WOCL. It therefore seemed worthwhile to combine the two FDP types. Table 44 shows that the main predictors in the simple logistic regression analysis (model 1) became FDP type (late plus night/non-disruptive), FDPWOCL, Sleep in 24 hours prior TOD, and FDP start time. In model 2 only the significant variables from model 1 were included. Sleep, start time and Function became significant predictors. In model 3, sleep was removed to permit other variables to enter the regression. This yielded start time, FDP duration, and Function as significant predictors. It may be argued that FDPs that start at, for example, 04:00h could be considered early FDPs and not constitute night work in the traditional sense, since night work commonly does not permit sleep during the traditional night window for sleep. Thus, we modified our selection of FDPs to only include start times before 00:00h. This would exclude extremely early start hours.

Table 44 Results from multiple logistic regression of predictors vs high fatigue. Late finish plus night dataset. N = 372

	Model 1 *Unadjusted OR (CI)*	***p***	***Model 2*** *OR (CI) Non-sign and FDPLFN and WOCL not entered*	***p***	***Model 3*** *OR (CI) Sleep removed*	***p***	***Model 4*** *OR (CI) Model 3 Start before 24h N = 320*	***p***
FDPdur: FDP duration	1.11 (1.02;1.21)	.014	1.09 (0.98;1.21)	.112	1.11 (1.00;1.23)	,049	1.17 (1.04;1.32)	.008
FDPWOCL	3.08 (1.77;5.34)	.000	#		#		#	
FDPsectors: No. of sectors	1.18 (0.92;1.51)	.189						
AwakeTOD: Time awake prior TOD	1.03 (0.99;1.07)	.096	0.95 (0.90;1.00)	.070	1.00 (0.95;1.05)	.957	1.01 (0.96;1.02)	.796
Sleep24h: Sleep in 24h prior TOD	0.82 (0.74;0.91)	.000	0.86 (0.75;0.99)	.029	#			
StartH: FDP start time	1.07 (1.03;1.11)	.000	1.05 (1.01;1.10)	.020	1.06 (1.02;1.11)	.007	1.10 (1.03;1.17)	.006
Function (1 pilot/ 2 cabin)	1.97 (1.67;3.24)	.011	2.06 (1.17;3.63)	.012	2.16 (1.24;3.77)	.006	2.22 (1.21;4.08)	.010
Gender (1 male/ 2 female)	0.58 (0.33;1.01)	.053						
Age (in years)	1.02 (0.99;1.05	.269						
FDPsleep: Napping	0.97 (0.93;1.01)	.191						
TZ_WE: Time zones crossed WE	1.02 (0.92;1.14)	.709						
TZ_EW: Time zones crossed EW	0.98 (0.86;1.11)	.722						
FDPLFN: FDP late finish plus night	3.05 (1.74;3.53)	.000	#		#		#	

OR = Odds Ratio. CI = 95% Confidence Interval. # = removed variable. p = level of significance.

Model 4 shows that this approach yielded FDP duration, FDP start time and Function as significant predictors, just as in model 3. This made the effect of FDP duration clearer than in model 3 and that of Function weaker. Note that also in this analysis FDP duration became significant when sleep was removed. As in the previous analysis end time is not a suitable variable for inclusion.

Excluding the non-disruptive controls

If we exclude the daytime controls and only use late finish and night FDPs, the simple logistic regression shows four significant predictors of high fatigue: FDP end time (OR = 0.95 (CI = 0.91; 0.99), p = .031), FDPWOCL (OR = 2.88 (CI = 1.31; 6.32), p = .008), Sleep in 24 hours prior TOD (OR = 0.85 (CI = 0.85; 0.74), p = .014), and Function (OR = 2.20 (CI = 1.01; 4.70), p = .041). For these analyses N = 142, with minor variations between variables. However, the multiple logistic regression with all significant predictors entered has a collinearity problem between WOCL and end times. We interpret this to mean that, as before, end time, WOCL, and sleep compete for the same variance and prevent a stable result; i.e., they are not independent. We also interpret the results to mean that the predictors of high fatigue are end time, prior sleep, WOCL and Function (higher fatigue in cabin crew).

For the late finish and night FDPs we also used maximum KSS as outcome in the same analysis as above. This yielded a significant result for end times only (OR = 1.13 (1.04; 1.24), p = .007). Among those significant when using KSS at TOD as outcome, sleep in previous 24 hours was not significant (OR = 0.89 (0.78; 1.01) p = .063), neither was WOCL (OR = 2.00 (0.98; 4.14) p = .060), nor Function (OR = 1.94 (0.94; 4.0) p = .07). N = 140. The results for this analysis of maximum KSS values is weaker than that for KSS at TOD.

Plots of end times vs high sleepiness and other predictors

To illustrate the results in another way, we also plotted end times vs the proportion of high fatigue and FDP duration (Figure 17). The proportion of high fatigue at TOD increased from FDP end times of 19:00h to 23:00h, and were highest for morning FDP end times. Note that the morning hours had the highest probability of high fatigue, not the WOCL. Also note that evening and early night end times are dominated by late finish FDPs. The results also show that evening and night end times are characterized by short FDP durations, that is, the two variables interact (most likely because FTL limitations on night FDPs). As consequence it is very difficult to evaluate the simultaneous effect of end time and FDP duration on high sleepiness.

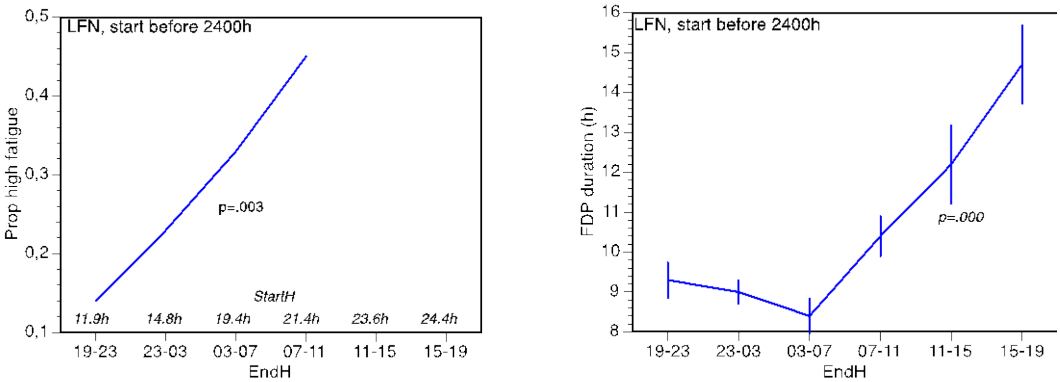


Figure 17 End times (EndH) plotted against probability of high fatigue and FDP duration. LFN = late finish plus night category

Furthermore, Figure 18 shows that sleep duration falls from the evening (mainly late finish FDPs) to afternoon FDPs (long duration night FDPs). Figure 18 also shows that evening and early night hours do not contain encroachment on the WOCL, whereas all remaining times involve such encroachment.

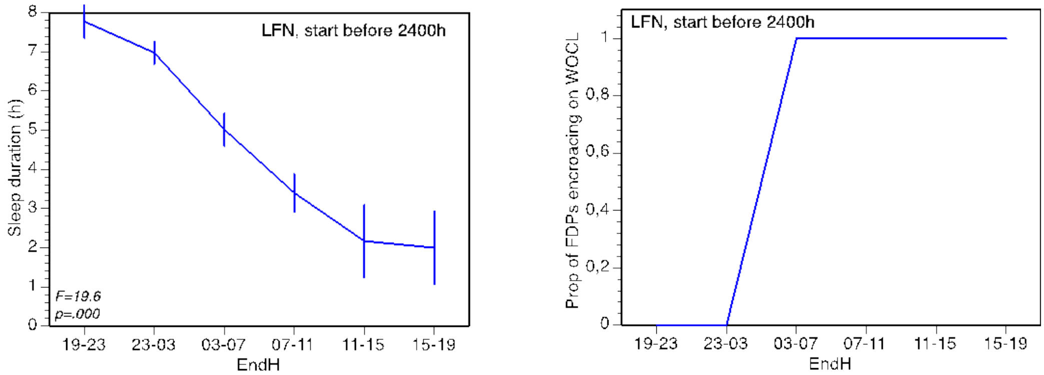


Figure 18 End times (EndH) plotted against prior sleep duration and probability of FDPs encroaching on the WOCL. LFN = late finish plus night category

A note on napping on the flight deck

The number naps on the flight deck in late finish and night FDPs is 32 (out of 144 FDPs = 22.4%). Their mean±SD duration was 19.6 ±7.8 minutes. Non-disruptive FDP controls had no naps. FDPs that encroach on the WOCL contained 19 naps (out of 92 FDPs = 20.7%). Their mean±SD duration was 19.9±1.9 minutes. We also selected those FDPs that encroached on the WOCL without starting in the WOCL, and found that OR and CI for minutes of napping vs probability of high fatigue was not significant (OR = 0.98 (CI = 0.85; 1.01), p = .085). N = 45.

A note on KSS TOD and MAX

In order to understand if there was any difference between KSS at TOD (during the last sector) and the maximum value (from TOD during any sector or from cruise phase during long-haul flights) we t-tested the difference. This resulted in mean±SD values for KSS TOD = 5.22±.1.95 and for KSS MAX = 5.51±1.97, with t = 3.66 and p = .000.

A comment on classification of FDPs as late finish or night

The results of the present analysis suggest that the subdivision of FDPs into late finish or night may be too narrow. There rather seems to be a continuum of lateness from early evening to late night/early morning that affects fatigue. Fatigue certainly occurs during the WOCL but even more so during the immediate post-WOCL hours, and to certain extent during the immediate pre-WOCL hours. Our results also indicate that flights starting during the WOCL do not seem to result in as high likelihood of fatigue as those that start before the WOCL. The former appears to represent a form of very early FDPs (discussed in the next on the variable very early start FDPs).

INTERIM CONCLUSION

The higher odds of high fatigue at TOD during late finish plus night FDPs, as compared to non-disruptive (daytime) FDPs, were mainly due to a later start time and a longer FDP duration.

When late finish plus night FDPs were considered alone, an earlier end time, shorter prior sleep, and encroachment on the WOCL are the FDP-related characteristics that predicted the occurrence of high fatigue at TOD.

Analyses of the variable very early start FDPs

The complement to the analyses of late finish plus night FDPs is a corresponding analysis of both early and very early starts; i.e., FDPs with start times between 03:00h and 06:59h. The results in Table 45 show that Sleep in 24 hours prior TOD and FDP type (very early vs non-disruptive FDPs) showed significant ORs in the simple logistic regression analysis. High fatigue increased with less sleep and with earlier start time. When both were entered in model 2 only sleep remained significant. The results are rather weak, but our interpretation is that mainly sleep loss is related to high fatigue, but that earlier start time contributes to short sleep. The correlation between the two variables was r = -0.33 (p < .001); i.e., short sleep was associated with early starts.

Table 45 Results from multiple logistic regression of predictors vs high fatigue. Early plus very early starts (between 03:00h and 06:59h) together with non-disruptive controls. N = 301

	Model 1 *Unadjusted OR (CI)*	***p***	***Model 2*** *OR (CI) Only significant predictors*	***p***
FDPdur: FDP duration	1.09 (0.98;1.20)	.105		
FDPWOCL	1.77 (0.87;3.60)	.115		
FDPsectors: No. of sectors	1.15 (0.88;1.50)	.359		
AwakeTOD: Time awake prior TOD	1.03 (0.98;1.08)	.285		
EndH: FDP end time	0.98 (0.93;1.04)	.524		
Sleep24h: Sleep in 24h prior TOD	0,85 (0.75;0.96)	.008	0.87 (0.77;0.99)	.042
StartH: FDP start time	0.93 (0.86;0.99)	.042	0.95 (0.88;1.03)	.197
Function (1 pilot/2 cabin)	1.34 (0.75;2.40)	.325		
Gender (1 male/2 female)	0.90 (0.47;1.71)	.749		
Age (in years)	0.98 (0.95;1.01)	.201		
FDPsleep: Napping during FDP	0.98 (0.93;1.03)	.455		
TZ_WE: Time zones crossed from West to East	1.00 (0.86;1.15)	.956		
TZ_EW: Time zones crossed from East to West	1.01 (0.88;1.17)	.871		
FDPES37: FDP early start between 3h and 7h	1.81 (1.03;3.18)	.039	#	

OR = Odds Ratio. CI = Confidence Interval. # = removed variable. p = level of significance.

<u>Excluding the non-disruptive controls</u>

We also analysed the group with early and very early start FDPs separately (without the non-disruptive controls). Only sleep was significant in the simple logistic regression analysis (OR = 0.79 (0.647; 0.970), p = .024).

No significant results were obtained when maximum KSS was used as outcome.

INTERIM CONCLUSION

The somewhat higher odds of increased fatigue at TOD during very early start FDPs, as compared to non-disruptive (daytime) FDPs, were mainly due to shorter prior sleep, which in turn was due to earlier FDP start time.

When very early start FDPs were considered alone only the amount of sleep in the past 24 hours significantly predicted which of these FDPs involve high fatigue at TOD. Longer sleep, which was associated with later FDP start time, was associated with a lower likelihood of high fatigue.

Additional analyses involving FDP duration

Comparison between long- and short-night FDPs by restricting FDP start time

As with the late finish plus night FDPs also the analysis of short-/long-night FDPs and high fatigue contain FDPs that start in the WOCL, for example at 03:00h or 04:00h. We repeated the analysis of short-/long-night FDPs and high fatigue, by restricting inclusions of those FDPs that started before 00:00h (and encroached on the WOCL). As in the previous analysis, the cut-off used was 10h duration.

The logistic regression analysis with short/long as predictor yielded OR = 1.23 (0.61; 2.44), N = 141. Testing other cut-offs did not change the results. The analysis confirms the previous results of no difference in high fatigue between long- and short-night FDPs,

Other observations on FDP duration and high fatigue

FDP duration did not show any significant association with high fatigue in the previous analyses, despite the fact that this variable is a central focus in FTL. A possible explanation may be that a potential association could be confounded by the large range in end times for night FDPs. This variation includes end times during the circadian low, which should increase high fatigue, as well as end times during the circadian upswing during late morning/noon, which would counteract such effects. Hence it might be of interest to analyse the association between FDP duration and high fatigue separately for night FDPs that end during the night and for those that end during the day.

We set the cut-off for night and daytime end times to 06:00h and carried out a logistic regression analysis. For end times < 06:00h with FDP duration as a predictor we obtained OR = 1.04 (CI = 0.86; 1.27), p = .673, N = 77, mean±SD = 8.36±2.46, range = 03:25h - 15:83h. For end times ≥ 06:00h the analysis for FDP duration yielded OR = 0.85 (CI = 0.63; 1.13), p = .268, N = 58, mean±SD = 10.97±1.86, range = 6.75 - 14.67. Changes of the 06:00h cut-off did not produce any different results. The results indicate that the association between FDP duration and high fatigue was not significant in the new analyses.

With respect to the effect of FDP duration one needs to consider that only two FDPs were found in the ≤ 4h category (out of 136 night FDPs). This clearly limited the ability to evaluate the effect of FDP duration across its whole continuum during night FDPs. Another characteristic of night FDPs that we found in the analysis of late finish plus night FDPs, is that those ending during the night were shorter than those ending later, apparently due to the FTL regulations. This creates a bias in the analyses that tend to reduce any effect of long FDP durations on high fatigue. In the present data, night FDPs of > 12h duration end at 8.42±2.92h; of 10 - 11.9h duration end at 6.90±2.43h; of 4 - 9.9h duration end at 4.40±1.6h; while the interval of 1 - 3.9h duration contains too few participants (N = 2). A further complication is that long duration FDPs co-occur with the WOCL, considerable sleep loss, and long time awake. These are extremely strong determinants of fatigue and may easily overshadow any effect of FDP duration.

The lack of clear effect of FDP duration on high fatigue is counterintuitive, but agrees with other studies that have studied this phenomenon while sleep duration is controlled (Gander et al. 2013, 2014, 2015) and in cabin crew (van den Berg et al.

2015). When sleep is not controlled, FDP duration appears a significant factor, as in the study by Powell et al. (2008). This does not mean that FDP duration is of no importance, instead, long exposure at high fatigue will likely be experienced as a considerable burden, even if fatigue does not increase above 'high' levels (e.g. KSS ≥ 7).

In a final attempt to investigate FDP duration in relation to fatigue, we also analysed how FDP duration and all other predictors were associated with fatigue *during daytime FDPs*, using logistic regression. We have approached this issue in prior analyses when the different FDP types of interest were removed, but the results were not significant. Possibly the number of available FDPs was too low. Here we selected all available first daytime FDPs for all participants, with daytime referring to the time from 07:00h to 23:00h. This yielded 244 individuals who had an FDP in that interval. The only significant predictors of high fatigue became FDP duration (OR = 1.24, CI = 1.09 - 1.41, p = .001) and End time (OR = 1.13, CI = 1.00 - 1.28, p = .045). No other variables were significant. Entering the two predictors into a multiple logistic regression, we obtained a reduced, but still significant OR = 1.21 (CI = 1.04 - 1.39, p = .012) for FDP duration and a non-significant result for End time. The mean±SD of FDP duration was 7.15±3.00h, range 0.75 - 13.8h. The range of variation is considerable, which is in contrast to the lack of short FDPs in the night FDPs. This difference might contribute to the difference in impact of FDP duration during night and daytime FDPs (apart from the contributions of circadian low and sleep loss).

A suggestion for alternative FDP categories

The above-presented results were mainly based on FDPs as they are currently classified in the FTL regulations. The analyses suggest that the current way of classifying FDPs may not be optimal. Here we propose an alternative way of classifying FDPs based on their start and end times. This new classification is closely linked the three-process model of alertness, which is an established and science-based model to predict sleep and fatigue (Åkerstedt & Folkard, 1997).

One of the main revisions in the alternative classification is to have two early start FDP categories: deep early (start between 02:00h and 04:59h) and early (start between 05:00h and 06:59h) FDPs. In the current FTL-based classification, all FDPs starting between 02:00h and 04:59h are considered as night FDPs, even though the crew usually obtain at least some night sleep just prior to them. This phenomenon makes the FDPs that start within this time window quite different from those that start, for example, before midnight and end during or after the WOCL. The proposal of having these two categories for early start FDPs is based on the observation that the earlier a duty starts, the less prior sleep can be obtained and the stronger is the fatiguing effect of the circadian process especially during the first 1 to 3 hours on duty.

Another suggestion is to have two categories for night FDPs as well. The night category (end time between 02:00h and 05:59h) includes FDPs with TOD - blocks-on phase inside the WOCL whereas the deep night category (end time between 06:00h and later) includes FDPs with the TOD - blocks-on phase after the WOCL. These two night FDP categories also differ in terms of how completely they cover the WOCL, only the latter covering the window completely.

Figure 19 shows the occurrence-probability point estimate of high fatigue at TOD in each FDP category as compared to non-disruptive FDPs (start time at 07:00h or later and end time earlier than 23:00h). Each FDP of interest showed an increased tendency towards high likelihood of fatigue. The tendency was especially discernible during late finish, night, and deep night FDPs, and of these, the tendency was the most pronounced in the deep night category. It is also worth mentioning that the deep early and early start categories showed practically no difference.

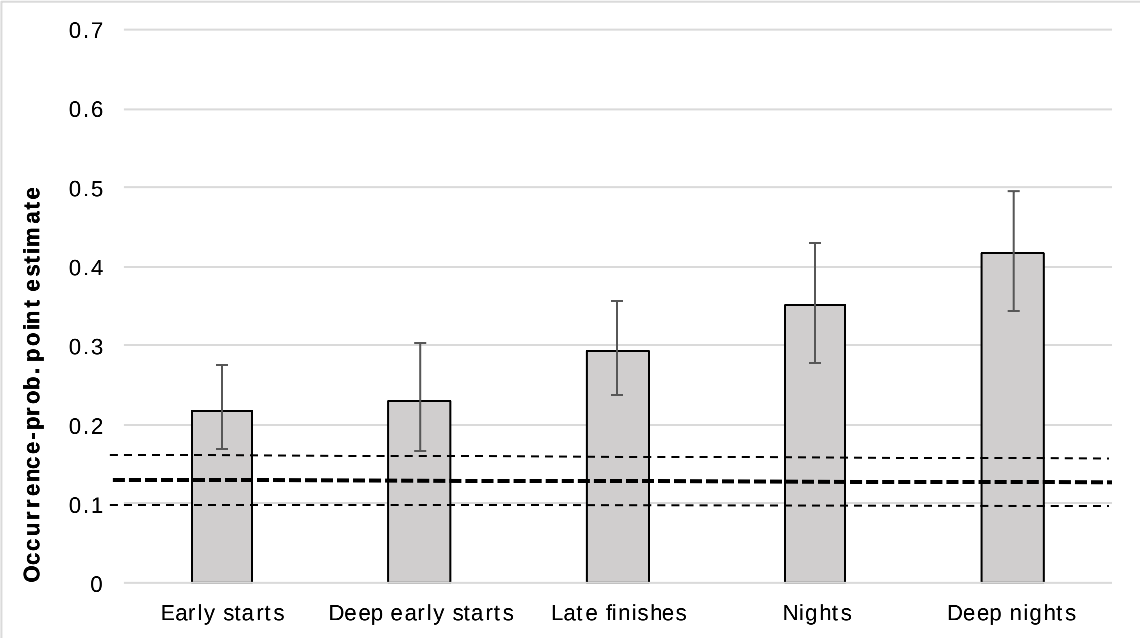


Figure 19 The occurrence-probability point estimate of high fatigue at TOD in each FDP category of interest. The vertical lines denote a 95% CI. The thick dotted horizontal line denotes the non-disruptive (daytime) FDP category (all FDPs with start time ≥ 07:00h and end time ≤ 22:59h) and the thin dotted lines a 95% CI

Table 46 shows the results of a logistic regression analysis for each FDP category of interest. In each of these categories, the OR of reporting high fatigue at TOD was increased as compared to daytime FDPs. The most pronounced increase was found in the deep night category. In this category, the OR was about twice as high as in the other four categories.

Table 46 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.

The main outcome in the analyses has been the occurrence of high fatigue at TOD during the final sector. However, we also analysed the data by taking into account the KSS ratings at TOD during all the other sectors, if such existed, and also during the cruise phase (Figure 20). The results were very similar to those based on fatigue at TOD during the last sector, but consistently somewhat higher. The major clear difference between these two outcomes can be seen in the deep night category. In this category, the occurrence-probability point estimate of high fatigue showed an 11-point increase (0.42 vs 0.53) as compared to the result based on a rating at TOD only.

We do not have data to show exactly the reason for this, but our tentative explanation lies in two observations. These FDPs were mainly long-haul and the cruise phase was extended (see below). Both of these characteristics can be suspected to increase the likelihood of high ratings given during cruise phase, which is a period of reduced activity compared to TOD.

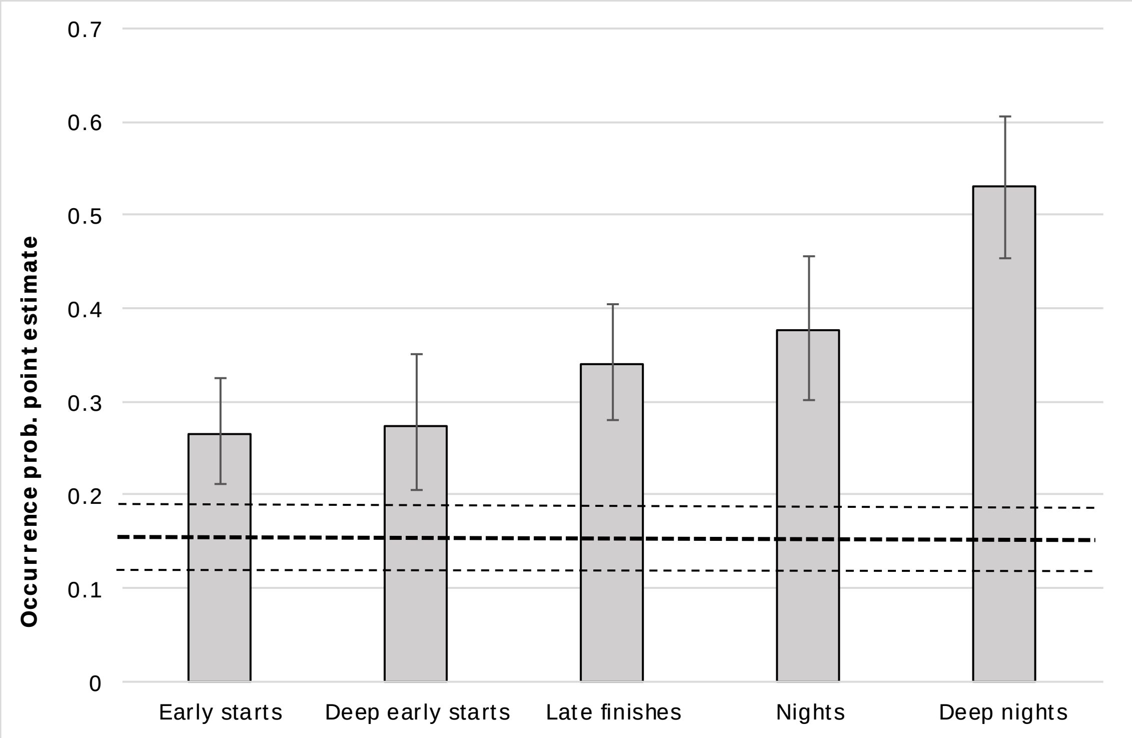


Figure 20 The occurrence-probability point estimate of high fatigue at either cruise phase or TOD or both in each FDP category of interest. The vertical lines denote a 95% CI. The thick dotted horizontal line denotes the non-disruptive (daytime) FDP category and the thin dotted lines a 95% CI

Figure 21 shows the amount of prior sleep and wake in each FDP category. The main observation was a decrease in prior sleep and an increase in prior wake in connection with late finish, night, and deep night FDPs. This tendency seemed to be the most pronounced in the deep night category. The ratio between prior sleep and wake was, on average, 0.5 or greater (at least 1 hour of sleep for 2 hours of wakefulness) for the non-disruptive, early start, and deep early start categories. For late finish FDPs, the ratio approximated this cut-off point (0.45). For the two night categories, the ratio was clearly below 0.5, being only 0.22 for deep night FDPs.

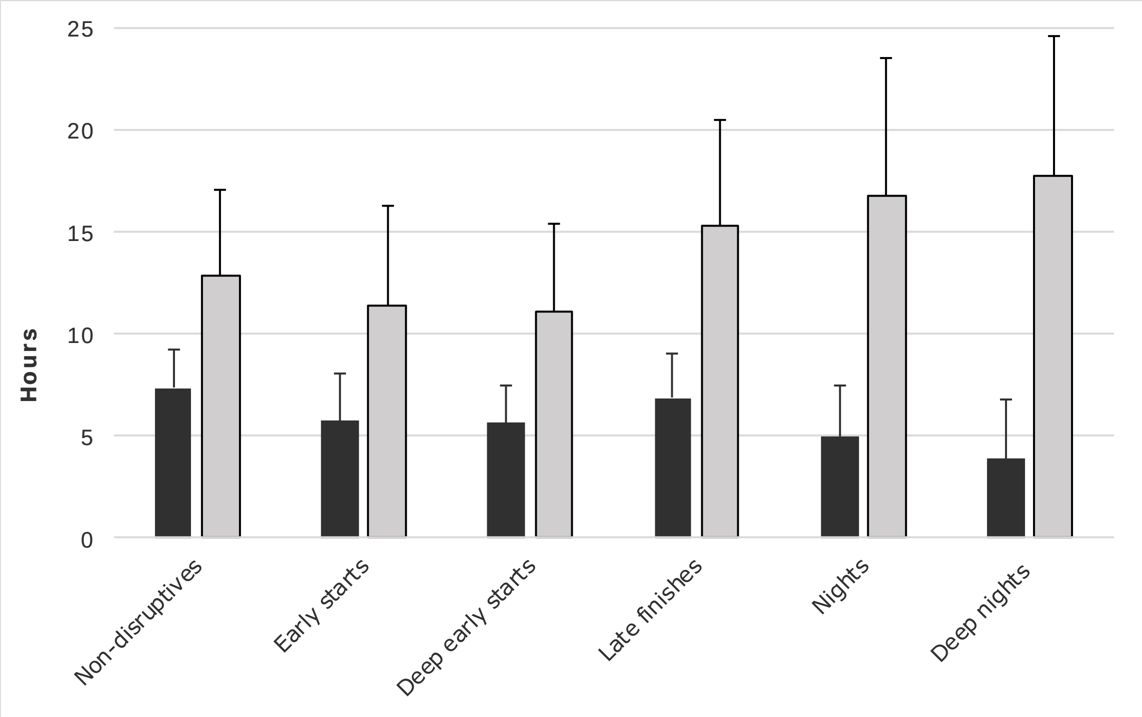


Figure 21 Mean (SD) amount of prior sleep (black bars) and wake (grey bars) in the FDP categories of interest and their non-disruptive (daytime) FDP category

Figure 22 and Figure 23 show the duration and number of sectors for each FDP category of interest and their non-disruptive FDPs. The main finding turned out to be a very long duration in the deep night category. This category contains FDPs that ended during daytime; the previously observed short duration for night FDPs focused on those ending during the night or early morning.

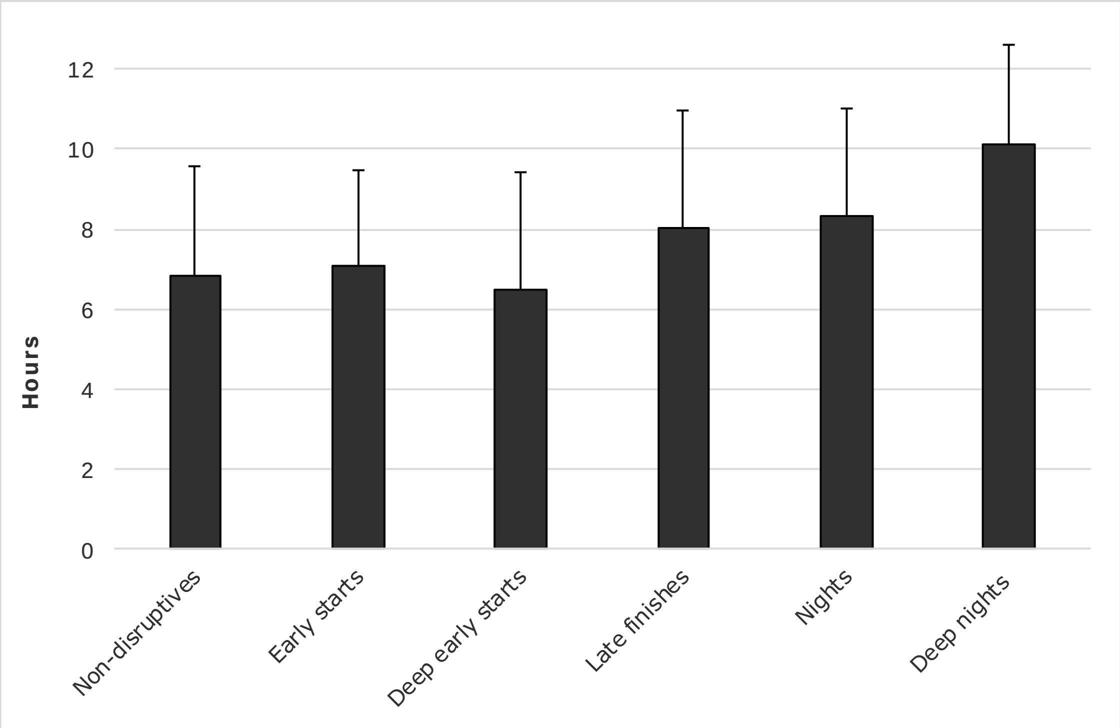


Figure 22 Mean (SD) FDP duration in the FDP categories of interest and their non-disruptive (daytime) FDP category

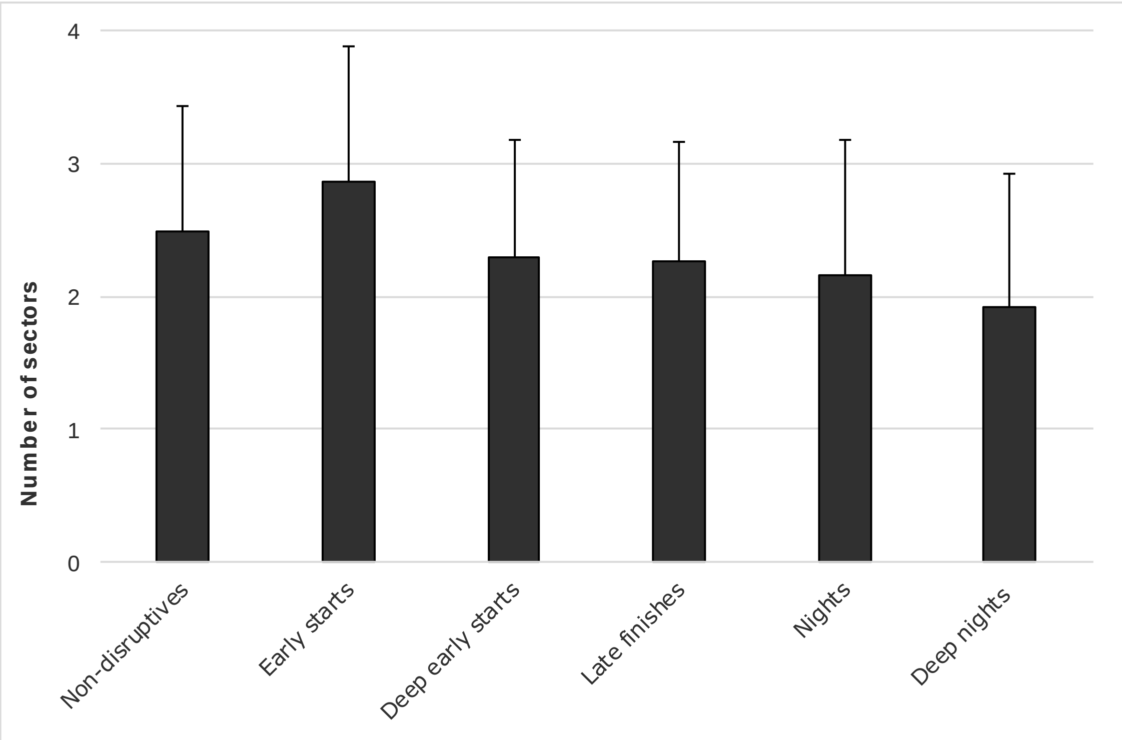


Figure 23 Mean (SD) number of sectors in the FDP categories of interest and their non-disruptive (daytime) FDP category

Summary of the results

These results are exclusively based on the current field data 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	***% of FDPs*** *Entire data*[1]	***OR*** *Btw-Ss data*[2]	***Main results on high fatigue[3] at TOD and its predictors***
Night duties (> 10 h)	41 (vs 24)	1.06[ns]	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 night FDPs (> 10h).
Nights	34 (vs 24)	3.21***	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 form evening to night, a new FDP category called late finish plus night was formed (start time before 00:00h). Higher 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]. Higher probability of high fatigue at TOD during these FDPs compared to during daytime FDPs was explained by earlier FDP start time and shorter prior sleep. When these FDPs were analysed alone, only shorter prior sleep explained the occurrence of high fatigue at TOD.
			An alternative way of classifying FDPs was suggested. When applying this classification, the probability of high fatigue at TOD was found to be similar for deep early (start time 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	19 (vs 24)	2.02*	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	31 (vs 24)	3.77***	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.

OR Odds Ratio. Ns = non-significant. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

[1] Percentage of FDPs with high fatigue at TOD (KSS 7 - 9) during the FDPs of interest (vs all FDPs) in the entire data.

[2] OR based on the occurrence of KSS 7 - 9 at TOD during the FDPs of interest vs daytime FDPs in the between-subjects data (btw-Ss), except for long-night FDPs (> 10h) that are compared to short-night FDPs (≤ 10h).

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

[4] FDPs starting between 03:00h and 04:59h are considered night FDPs in the current FTL, whereas those starting between 05:00h and 06:59h are not.

[5] 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	***% of FDPs*** *Entire data*[1]	***OR*** *Wth-Ss data*[2]	***Main results on high fatigue*[3] *at TOD***
Consecutive early starts	7 (vs 24)	0.50[ns]	In the data collected, fatigue levels at TOD were similar for the first and second consecutive early start FDPs.
Consecutive late finishes	29 (vs 24)	-	In the data collected, it seemed that the probability of high fatigue at TOD was similar for the first and second consecutive late start FDPs.
Consecutive nights	35 (vs 24)	1.30[ns]	In the data collected, fatigue levels at TOD were similar for the first and second consecutive night FDP.
Mix	25 (vs 24)	-	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 baseline dataset, containing all FDPs collected.

OR Odds Ratio. Ns = non-significant.
[1] Percentage of FDPs with high fatigue at TOD (KSS 7 - 9) during the FDPs of interest (vs all FDPs) in the entire data.
[2] OR based on the occurrence of KSS 7 - 9 at TOD during the first FDP of interest vs the second FDP of interest in a row in the within-subjects data (wth-Ss).
[3] A high level of fatigue was defined by KSS scores equal or greater than 7 (= sleepy, but no effort to keep awake).

Measurement tools

The different measurement tools were compared to determine whether or not different tools provided the same results.

Compare actigraph with sleep log

Research questions

Is there a significant difference between sleep duration determined with the actigraph and sleep log? And is there a significant correlation between sleep duration determined with the actigraph and sleep log?

Participants

When crew members registered to participate in the data collection, they were asked whether they were willing to wear an actigraph. The actigraphs were sent out to the volunteers based on their availability. A total of 91 actigraphs were sent to participants. We were able to retrieve data from 68 actigraphs. The remaining 23 actigraphs were not worn, were lost, or were worn outside the 30-day activation period of the actigraph.

Method[14]

The actigraph data was collected and processed for further use. We used the following criteria for this:

- Every 24 hours that a participant had continuously worn the actigraph, with exception of time for shower or bath, were analysed;
- Sleep periods of longer than 25 minutes were identified and selected within the 24-hour-periods;
- Interruptions in sleep of less than 10 minutes were ignored during a longer period of sleep and were also considered sleep time;
- We considered a registered period of longer than 2 hours of complete inactivity as missing data; and
- The sleep time reported by the actigraph software was considered. Other characterisations from the software were not used in the analysis.

The sleep log data was collected and processed accordingly. Reported sleep per participant in the 24-hour periods overlapping the periods registered from the actigraph was included in the analysis; i.e., the combination of sleep log and actigraphy data.

Results

A total of 241 (excluding six outliers) 24-hour-periods collected by 19 participants (i.e., 5% of all 381 participants) were analysed. A two-tailed Wilcoxon Signed-Ranks test was executed to determine the difference between sleep duration determined with the actigraph versus the combination of sleep log and actigraph. The results indicated no statistical significant difference between the actigraph versus the combination of sleep log and actigraph ($Z = -.325$, $p = .745$).

A Pearson correlation test was run to determine the relationship between sleep duration determined with the actigraph versus the combination of sleep log and actigraph. There was a strong, positive correlation between actigraph and the combination of sleep log and actigraph, which was statistically significant ($r = .888$, $p < .001$). In addition to the Pearson test, a Kendall Tau-b and a Spearman rank-order

[14] See D2.2 (Definition of the Data Collection Process) for a detailed description of the method used.

correlation test were run. Both tests showed strong, positive correlations between actigraph and the combination of sleep log and actigraph data, which were statistically significant (respectively τ_b = .655, p < .001 and (r_s = .828, p < .001).

Table 47 Descriptive statistics sleep duration outcome measures

Sleep duration outcome measure	***Sleep periods*** *No.*	***Minimum*** *Minutes*	***Maximum*** *Minutes*	***Mean*** *Minutes*	***SD***
Sleep log	241	25	1135	390.36	169.284
Actigraph	241	25	1136	397.88	182.284

SD = Standard Deviation.

Figure 24 shows a scatterplot for sleep duration in actigraph and the combination of sleep log and actigraph with an R^2 value of .789. A linear regression line has also been fitted resulting in the following model for the association between the sleep duration in actigraph and combination of sleep log and actigraph (independent variable X (sleep duration of sleep log), dependent variable Y (sleep duration of actigraph)): $Y = 25.778 + .953 * X$. The model states that sleep duration measured by the actigraph can be obtained from sleep duration in the sleep log by multiplying the latter by a factor 0.953 and adding 25.778 (all times in minutes).

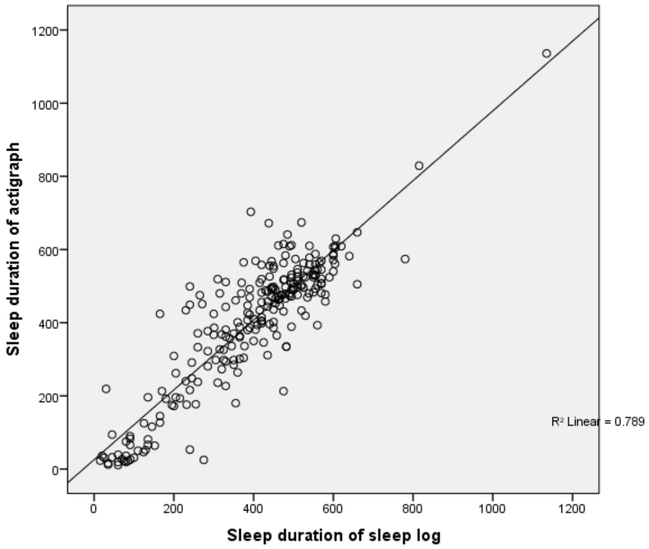


Figure 24 Scatterplot of sleep duration in sleep log (in min) – sleep duration in actigraph (in min)

<u>Conclusion</u>

We found no statistical difference between the actigraph versus combination of sleep log and actigraph data. Also, we identified a statistical significant correlation between the actigraph and combination of sleep log and actigraph data. Therefore, we concluded that for the current dataset the self-reported sleep log in the data collection app could be considered an accurate measure of sleep duration as compared to the combination of logs and objective actigraph data.

We only made the comparison of the sleep duration measured with the actigraph and sleep log after we had received data from 25 participants, of which 19 uploads were considered useful. We performed this preliminary analysis because we needed confirmation that the self-reported sleep could be considered an accurate measure to determine sleep duration before the data collection had ended. Otherwise, in case of a significant difference between the actigraph and sleep log data, the manner in which we measure sleep duration within the FTL data collection had to be re-considered. The preliminary analysis provided us this confirmation. Therefore, no further action was undertaken with regard to the actigraph and sleep log comparison. Hence, for the rest of the data analysis only the sleep duration resulting from the sleep log is used.

Compare KSS and SP

Research question
Are there significant correlations between the KSS and SP measures?

Participants
The entire crew population as described in the section 'Description of crew member data' was used for this analysis.

Method
The following outcome measures were compared:
· KSS TOD – SP TOD.

Results
We ran a Pearson product-moment correlation to determine the relationship between the different outcome measures.

There was a strong, positive correlation between KSS TOD and SP TOD (r = .868, N = 1632, $p < .001$).

A linear regression line was also fitted resulting in the following model for the association between the different pairs of measures (independent variable X (KSS TOD), dependent variable Y (SP TOD)): $Y = 1.118 + .566 * X$. The model states that SP TOD can be obtained from KSS TOD by multiplying the latter by a factor 0.566 and adding 1.118.

Table 48 Descriptive statistics outcome measures

Outcome measure	***N***	***Minimum***	***Maximum***	***Mean***	***SD***
KSS TOD (*1-9*)	1641	1	9	4.93	1.933
SP TOD (*1-7*)	1643	1	7	3.90	1.258

SD = Standard Deviation.

Figure 25 shows the scatterplot for the pair of outcome measures with an R^2 value of .753.

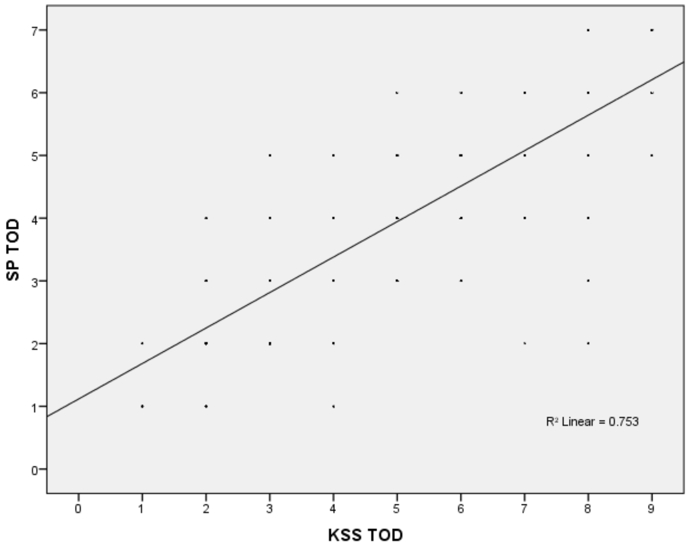


Figure 25 Scatterplot KSS TOD - SP TOD

Conclusion
Based on the results presented we concluded that the KSS TOD and SP TOD measures were strongly correlated. This implied that although the measures were not the same we could expect similar results from both measures. Therefore we decided to primarily focus the data analysis of FDP1 (Night duties of more than 10 hours) and FDP2 (Disruptive schedules) on only one measure instead of the two.

In line with the approach defined in D2.2 (Definition of the Data Collection Process), the primary data analyses were performed using the KSS TOD measure.

Compare mental effort with FDP sectors and number of contributing factors

Research question
Is there a significant correlation between Mental effort and FDP sectors? And is there a significant correlation between Mental effort and No. of contributing factors to workload?

Participants
The entire crew population as described in the section 'Description of crew member data' was used for this analysis.

Method[15]
Mental effort was considered by means of a subjective rating scale; that is, the level of mental effort experienced by the participant in the particular duty period was rated by moving a slide bar. The bar runs from 'almost no effort' to 'extreme effort' and is based on the rating scale mental effort (RSME) (Zijlstra, 1993).

The following measures were compared:
- Mental effort – FDP sectors (no. of sectors flown in FDP); and
- Mental effort – No. of contributing factors.

[15] See D2.2 (Definition of the Data Collection Process) for a detailed description of the method used.

Results

We ran a Pearson product-moment correlation to determine the relationship between the different measures.

Table 49 Descriptive statistics outcome measures

Outcome measure	***N***	***Minimum***	***Maximum***	***Mean***	***SD***
Mental effort (*0-150*)	2222	0	150	22.28	31.523
FDP sectors (*No.*)	2252	1	5	2.17	1.046
Contributing factors (*No.*)	2295	0	13	.63	1.641

SD = Standard Deviation.

There was a weak, positive correlation ($r < .05$) between Mental effort and FDP sectors ($r = .112$, $N = 779$, $p < .001$. And there was a strong, positive correlation between Mental effort and No. of contributing factors ($r = .689$, $N = 996$, $p < .001$).

Splitting the dataset into different Function groups (i.e., pilot and cabin crew) hardly had an impact on the strength and significance of the correlation between Mental effort and No. of contributing factors (for both pilot and cabin crew). However, for cabin crew the correlation between Mental effort and FDP sectors weakened ($r = .033$, $N = 254$) and was not significant anymore.

Splitting the dataset into different FDP types (FDP1 Night duties > 10h and FDP2 Disruptive schedules) did again not have any impact on the strength and significance of the correlation between Mental effort and No. of contributing factors (for both FDP types). However, for FDP1 the correlation between Mental effort and FDP sectors strengthened ($r = .234$, $N = 103$, $p < .05$); and for FDP2 the correlation between Mental effort and FDP sectors weakened ($r = .085$, $N = 386$) and was not significant anymore.

Splitting up both FDP types into the two Function groups showed for FDP1 an increase in strength of the correlation between Mental effort and FDP sectors ($r = .512$, $N = 23$, $p < .05$) for the cabin crew. For FDP1/pilot and FDP2/pilot/cabin crew correlations between Mental effort and FDP sectors were weakened and non-significant.

Conclusion

Based on the results presented we concluded that Mental effort and No. of contributing factors were strongly correlated. Note that contributing factors could only be administered in case participants scored 50 (somewhat effortful) or higher on the rating scale.

Splitting the dataset into different Function groups and FDP type did not impact this. This was different for the correlation between the Mental effort and FDP sectors measures. It turned out that there was a moderate correlation between Mental effort and FDP sectors for cabin crew in FDP1 (Night duties of more than 10 hours) only.

Chapter 4: 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).

The eastern region seemed to be overrepresented in the dataset with six airlines participating; however these six were smaller airlines as compared to the other airlines.

Data was collected by 381 crew members[16] 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[17]. A 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 estimated required sample sizes, as established in D2.2 (Definition of the Data Collection Process), were achieved for FDP1 (Night duties of more than 10 hours). This was not the case for all types of consecutive FDP2 (Disruptive schedules). The data collection resulted in sample sizes below the estimation for the consecutive late finishes and to a lesser extent for consecutive early starts. One explanation for this could be that the selection of airlines did not include enough airlines that frequently operate multiple consecutive late finishes and early starts. Another explanation could be that crew member participation was limited for those airlines operating these specific consecutive disruptive schedules. As the resulting dataset is completely anonymized (for airline and crew), we could not determine if either of the two or both explanations was/were correct. For the gathered FDPs within the current sample of airlines and crew we concluded that the occurrence of consecutive late finishes and early starts was less than for the consecutive nights and mixes of disruptive schedules.

In the sample size estimations outlined in D2.2 we included a subdivision by Function of the crew member; i.e., pilots and cabin crew. The provisions concerning flight and duty time limitations and rest requirements contained in Annexes II and III of Commission Regulation (EU) No. 965/2012 only (or limited) includes such a

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

subdivision in ORO.FTL.205.3(c) Reporting Time. Therefore, our primary focus in the data analysis was on all crew members without the subdivision by Function, but with Function as an independent variable.

For each duty in the dataset we checked whether the duty matched the criteria for FDP1 (Night duties of more than 10 hours)[18] or FDP2 (Disruptive schedules)[19]. The times used in these criteria were applied strictly to determine the final FDP1 and FDP2 dataset. This means that duties that started at 06:00h for example (in case of early type airlines) were not included in the FDP2 set while duties that started at 05:59h were included. This strict application of the criteria clearly had an impact on the resulting sample size for both FDPs.

The required sample sizes were estimates and were largely dependent on the expected effect size. Sallinen et al. (2017) is one of the few studies that reported the percentage of night duties where a score of 8 or 9 was measured on the KSS. Therefore, these percentages were used as an input for estimating the expected effect sizes. Within our study we were looking at KSS scores of 7, 8 or 9, implying a higher expected effect size and therefore a smaller required sample size. Also, the expected effect size for the consecutive disruptive schedules was unknown beforehand as no studies were found reporting this.

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

[18] Duties of more than 10 hours at the less favourable time of day. This refers to operations that encroach (part of) the night (the period between 02:00h and 04:59h).

[19] A crew member's roster which disrupts the sleep opportunity during the optimal sleep time window by comprising a duty or a combination of duties which encroach, start or finish, during any portion of the day or of the night where a crew member is acclimatised. A schedule may be disruptive due to early starts, late finishes or night duties (e.g. 'early type' of disruptive schedule means: for 'early start' a duty period starting in the period between 05:00h and 05:59h in the time zone to which a crew member is acclimatised; and for 'late finish' a duty period finishing in the period between 23:00h and 01:59h; 'late type' of disruptive schedule means for 'early start' a duty period starting in the period between 05:00h and 06:59h in the time zone to which a crew member is acclimatised; and for 'late finish' a duty period finishing in the period between 00:00h and 01:59h.

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[20] 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 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 $\geq$ 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.

Next, we will discuss our results in light of pertinent previous research.

Fatigue during FDP1 (Night duties of more than 10 hours)

The results of fatigue during night duties of more than 10 hours were two-fold: the probability of high fatigue at the TOD phase was quite high, but long duration, in isolation, was not significant predictor determinant of fatigue.

The evidence for a high likelihood of high fatigue at TOD was solid. Of all measured night FDPs (> 10h) 41% involved KSS scores of 7 or higher. This was clearly a higher relative frequency as compared to that calculated across all FDPs measures (24%).

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

This figure means that more than every third night FDP (> 10h) involved high fatigue at TOD. The mean level of fatigue measured at TOD fell between the items 'neither alert nor sleepy' and 'some signs of sleepiness' of the KSS. In addition, the frequency of napping on the flight deck (ideally using a controlled rest procedure) was high (27% of night FDPs > 10h) indicating increased sleep pressure. All these results are quite comparable to those found in a recent, extensive field-based study on commercial airline pilots (Sallinen et al. 2017).

When interpreting our finding of night duties of more than 10 hours, it is important to take into account that the long- and short-night FDPs (with 10h as a cut-off) differed in many respects, not only in duration. As compared to the short-night FDPs, on the long-night FDPs, the time spent awake was longer, sleep in the prior 24 hours shorter, and the time spent inside the WOCL longer. All these differences can be considered to increase the occurrence of high fatigue at TOD during the long-night FDPs as compared to the short-night FDPs. In addition, the number of time zones crossed was higher for the long-night FDPs, which may play a role especially in inbound FDPs. A higher number of time zones crossed during an inbound FDP means that the crew (i.e., acclimatised as we have excluded FDPs where crew members were in an unknown state of acclimatisation) have been exposed to a greater time difference at the destination, which, in turn, may affect the amount of prior sleep and the level of acclimatization in general. On the other hand, the number of sectors was lower, the amount of on-duty sleep greater, and the FDP end time more frequently outside the WOCL during the long-night FDPs than the short-night FDPs. These differences can be considered to decrease the occurrence of high fatigue at TOD during the long-night FDPs.

When comparing the long- and short-night FDPs, we made an attempt to control for these differences in order to see the main effect of FDP duration. This attempt did not, however, reveal a significant difference in the probability of high fatigue at TOD between the two types of night FDPs. This result suggests that the role of FDP duration, in isolation, is more limited during night FDPs as compared to non-disruptive and late finish FDPs that showed the main effect of FDP duration in our data. This result can be explained by the fact that high fatigue during night FDPs is mostly caused by the unfavourable time of the day and a reduced sleep-wake ratio, which easily masks the influences of other potential determinants such as FDP duration.

In light of previous studies our result is not surprising. A series of studies by Gander et al. (Gander et al. 2013; Gander et al. 2014; Gander et al. 2015) found that FDP duration was not a significant independent predictor of high fatigue at TOD on night-time transmeridian flights. A study by Powell at al. (2007) on short haul and mainly daytime and late finish FDPs, however, reported the opposite result. Together these studies suggest that the effect of FDP duration is easier to observe during daytime than night-time FDPs, which is in line with the results of the present study.

Our results suggest that high fatigue at TOD during night duties of more than 10 hours may be difficult to effectively control by just adjusting FDP duration. There are multiple other more influential determinants (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. 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.

Fatigue during FDP2 (Disruptive schedules)

We had two approaches to study disruptive schedules. First, we had all FDP types typical of these schedules (early starts, late finishes, and nights) pooled into the same category. This category consists of a heterogeneous group of FDPs in terms of start and end times and thus also in terms of the physiological mechanisms underlying fatigue. In spite of this heterogeneity, they all share one mechanism underlying high fatigue: they all overlap the usual 7 to 8 hour nocturnal sleep period.

Our main findings of this heterogeneous category were the following: i) 30% of all FDPs involved high fatigue at TOD; ii) the probability of occurrence of high fatigue was 1.26 times higher as compared to a baseline condition including all FDPs measured.

Next, we will discuss our results separately in each FDP type. It is worthwhile to keep in mind that our field data did not cover all FDP scenarios an operator could have under the current FTL. This especially holds for the scenarios that are extreme in nature.

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 non-disruptive/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, as outside commercial aviation, these duties are usually considered as early morning duties (Härmä et al. 2015). 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, to some extent, uncompelling can be explained by two factors: i) a relatively good ratio between the prior sleep and wake upon reaching the TOD; and ii) the occurrence of TOD well outside the WOCL.

Our finding of similar fatigue at TOD during the first and second early start FDPs in a row is somewhat unexpected, for two nights of sleep restriction is known to reduce alertness further than a single night (van Dongen et al. 2003). A possible explanation is that the amount of prior sleep was almost an hour longer prior to the second early start FDP than prior to the first one. A recent field-based study on commercial airline pilots suggests that already that degree of difference in prior sleep plays a role in on-duty fatigue (Sallinen et al. 2018).

A critical question here is what happens to fatigue if there are more than two early start FDPs in a row. Until now, most of the studies on consecutive work shifts have focused on night work (for a review Kecklund et al. 2017) even though some studies have focused on early morning shifts, too (Spencer et al. 1997; McGuffog et al. 2004). A large of body of laboratory-based research on sleep restriction, however, suggests that both cognitive performance and fatigue show a progressive deterioration over a sequence of more than two consecutive early start FDPs (Belenky et al. 2003; Haavisto et al. 2010; Lowe et al. 2017). The severity of the deterioration can be expected to be the greater the higher is the amount of lost sleep per night (Belenky et al. 2003; Lowe et al. 2017). In occupational settings, the degree of sleep loss while working morning shifts is the greater the earlier is shift start time (Ingre et al. 2008).

There are a few reservations to our evidence of increased fatigue at TOD during early start FDPs. First, 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 coincidence or closeness of the WOCL. Second, our data did not allow us to study long sequences of consecutive early morning FDPs (longer than two FDPs), as mentioned above. The same holds for work patterns where an early start FDP is preceded by a duty sequence that compromises sleep (e.g. quick transitions) or results in a delayed circadian rhythm (e.g. a sequence of late finish FDPs).

Late finishes

The evidence of increased fatigue during late finish FDPs was quite solid. First, 31% of all late finish FDPs involved high fatigue at TOD. Second, the probability of occurrence of high fatigue at TOD was 1.28 times higher for the late finish FDP as compared to the baseline condition including all FDPs measured. Third, the odds of high fatigue at TOD were 3.77 times higher during late finish FDPs than non-disruptive FDPs in the between-subjects data. In addition, the mean KSS rating at TOD was almost one unit higher during late finish FDPs than non-disruptive FDPs.

Long FDP duration was the only FDP-related characteristics to predict high fatigue at TOD during late finish FDPs compared to daytime FDPs. Longer FDP duration also predicted whether or not a late finish FDP involved high fatigue at TOD. Thus, long FDP duration proved more significant factor underlying fatigue at TOD in connection with late finish FDPs than early start or night FDPs.

Our data on two consecutive late finish FDPs was quite limited and thus no definitive conclusions can be made. However, the comparison between the mean KSS values of the first and second late finish FDPs does not support the accumulation hypothesis. This preliminary observation is somewhat unexpected, for two consecutive days with an extended wake period can be expected to compromise alertness more than a single day. A possible explanation for this preliminary finding could be that crew members are able to delay their circadian rhythms over two consecutive late finish FDPs, which, in turn, promotes their alertness during the latter part of the second late finish FDP. Generally speaking, it is easier to delay circadian rhythms than advance them because the internal clock period is typically longer than a 24 hour solar day.

It is also worthwhile to mention that our field-based data did not include all relevant FDP combinations where a late finish FDP may occur, such as having three or more late finish FDPs in a row. This shortage of data limits our possibilities to draw comprehensive conclusions on late finish FDPs as a part of a potential fatigue hotspot.

Based on the above-mentioned 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. Night duties are known to be the most fatiguing duties across industries (Sallinen et al. 2015).

Nights

We found five types of evidence of increased fatigue during night FDPs. First, 34% of all night FDPs involved high fatigue at TOD. Second, the probability of occurrence of high fatigue at TOD was 1.39 times higher for the night FDPs as compared to the baseline condition including all FDP measured. Third, the odds of high fatigue at TOD were 3.21 times higher for night FDPs as compared to non-disruptive FDPs. Fourth, the mean KSS rating at TOD was approximately 1 unit higher for the night FDPs than non-disruptive FDPs. Fifth, the relative frequency of naps on the flight deck was higher in comparison with the other FDPs of interest.

As already mentioned in connection with the night FDPs of more than 10 hours, our results of increased fatigue during night FDPs are well in accordance with previous studies and the physiological mechanisms underlying this phenomenon are well known. Also, not finding a significant result for the accumulation of fatigue over two consecutive night FDPs is in line with some field-based and laboratory-based studies showing that the first night shift, which is associated with a long period of wakefulness, can be even more fatiguing than the second night shift (Santhi et al. 2007; Pylkkönen et al. 2015). On the other hand, also opposite results have been reported and thus any definite conclusion is premature (for a review, see Kecklund et al. 2017).

When trying to assess an acceptable number of consecutive night FDPs, it is important to bear in mind that the accumulation of fatigue over a sequence of consecutive night FDPs is largely dependent on the degree of circadian adjustment to this sequence. In practice, the degree of adjustment is largely determined by the start times of consecutive night FDPs. A larger variability of start times within a sequence of consecutive night FDPs (start times can vary between 13:00h and 04:59h) makes it more difficult to reset the circadian pacemaker. In addition, the sequences of consecutive night FDPs that require one to advance bedtime (e.g. start at 04:00h) are probably more challenging than those that do the opposite because the internal clock period is typically longer than a 24 hour solar day, as mentioned above.

In our data, high fatigue at TOD during the night FDPs, as compared to daytime FDPs, was predicted by the time of day (i.e., encroachment on the WOCL) and a reduced amount of sleep in the past 24 hours. In addition, being a cabin crewmember (as opposed to a pilot) increased the odds of high fatigue. It is also noteworthy that the FDP characteristics examined did not predict the occurrence of high fatigue *within* the night FDP.

When the continuum from evening to night was covered with the late finish plus night FDP category (start time before 00:00h), encroachment on the WOCL and shorter prior sleep predicted both increased fatigue at TOD in comparison to daytime FDPs and the occurrence of high fatigue inside the late finish plus night category. This finding emphasises the role of two well-known determinants of fatigue, called the homeostatic and circadian processes, when attempting to identify fatigue hotspots and mitigate on-duty fatigue. This observation also gives reason to consider whether FPDs could be reclassified in a manner that is more firmly based on these processes than the current FTL-based classification.

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-categorized the night FDPs with start time between 02:00h and 04:59h as *deep early starts*. A reason for this revision was that crew are able to obtain some night sleep just before these extremely early start FDPs, as opposed to the other night FDPs. Another reason was that we wanted to see if deep early FDPs involve more fatigue than early FDPs (start time between 05:00h and 06:59h).

This re-categorisation did not reveal a sizable difference in high fatigue at TOD between the deep early starts and early starts. A possible explanation is a relative high sleep-wake ratio (0.5) prior to TOD during deep early starts. It is also worth noting that fatigue was not measured in the beginning of FDPs (e.g. at blocks-off or top of climb). This might be of importance, since the circadian effect can be assumed to peak in this flight phase during deep early starts.

We also divided night FDPs into *nights* (end time inside the WOCL) and *deep nights* (end time after the WOCL) and found an exceptionally high fatigue during the latter. A tentative explanation lies in three observations: the sleep-wake ratio was very low (0.22), the FDPs covered completely the WOCL, and FDP duration was long.

In summary, our results suggest that late finishes and nights 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 linked more closely to an established and science-based model used to predict fatigue (e.g. three-process model of alertness). This revision would probably allow for better management of fatigue.

Validation of the fatigue hotspots for the airline roster data

This section contains a comparison of the results from the field study with the results from the airline roster data and online survey[21].

FDP1 (Night duties of more than 10 hours)

With regard to the prevalence of high fatigue scores during FDPs longer than 10 hours that encroach (part of) the period between 02:00h and 04:59h we found similar results for the roster data and the data resulting from the aircrew data collection. High fatigue scores occurred in both datasets and the proportion of high fatigue for long duties encroaching on the night was higher than in the FDP Baseline datasets used. Also, the survey results indicated 'a long working day' (which is linked to long nights) and 'flying during hours when I would normally sleep' were selected as contributors to fatigue items by around 14% and 11% of aircrew.

The finding of the non-significant role of FDP duration corresponds with the results from the roster analysis, where we found hardly an effect on the odds of high fatigue at TOD.

FDP2 (Disruptive schedules)

The prevalence of high fatigue scores during disruptive flight duties showed – to a large extent – similar results for the roster data and the data derived from the aircrew study.

Results from both datasets aligned as it concerns nights (consecutive and not). In both sets we indicated relatively high prevalence for nights and the proportion of high fatigue scores was higher for the consecutives (except for the non-consecutive nights) than for the FDP Baseline sets. This was confirmed by a relatively high percentage of the fatigue item 'flying during hours when I would normally sleep' (11%) mentioned in the survey.

For the early start FDPs both datasets also aligned. However, here we found a low prevalence for consecutives and non-consecutives. This was not confirmed by the survey results, which showed a relatively *high* percentage of 'starting early' (12%) as a relevant fatigue item. This could be explained by the fact that the survey questions used the term 'fatigue' (broadly referring to physical fatigue, mental fatigue, and sleepiness), whereas the aircrew diary study primarily focussed on sleepiness (through the KSS measure).

Concerning late finishes we showed inconclusive results for the measures used in the roster data analysis. The difference in results was even bigger when looking at consecutive duties as compared to the non-consecutive ones. What we saw in both datasets (at least for one of the measures used in the roster data analysis) was a relatively high prevalence for late finishes and a proportion of high fatigue scores that was higher for the consecutive late finishes (again for one of the measures used in the roster data analysis) than for the FDP Baseline sets. This was confirmed by a relatively high percentage of the fatigue items 'late finishes' (7%) mentioned in the survey.

[21] As presented in D2.1 (Identification of Potential Fatigue Hotspots).

Early starts
The finding of a weak association between early starts and fatigue corresponds with the results from the roster analysis that showed that high fatigue scores did not occur more frequently in early starts compared to the baseline dataset (containing all FDPs). The finding of no-build-up of fatigue over two consecutive early starts corresponds with the results from the roster analyses where we found that high fatigue scores did not occur more frequently in consecutive early starts than in the baseline dataset.

Late finishes
The finding that for late finishes the level of fatigue was significantly higher than for non-disruptive FDPs partially corresponds with the results from the roster analysis, where one of the models produced results that showed higher fatigue for late finishes compared to the baseline dataset, while the other model produced results that did not show higher fatigue for late finishes compared to the baseline dataset.

The finding of no-build-up of fatigue over two consecutive late finishes does not seem to correspond with the results from the roster analysis, where we found an increase in percentage of high fatigue scores for the second late finish.

Nights
The finding that for night duties the level of fatigue was significantly higher than for non-disruptive FDPs corresponds with the results from the roster analysis where we found a significant result that high fatigue scores occurred more frequently in nights compared to the baseline dataset (containing all FDPs).

The finding of no-build-up of fatigue over two consecutive night duties does not seem to correspond with the results from the roster analysis, where we found an increase in percentage of high fatigue scores for the second night.

Chapter 5: 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.

The results of the assessment were grouped into three groups:
1. Recruitment and training of crew members;
2. Data collection tools; and
3. Measurement protocol.

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[22] 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[23].

Open invitations were sent to all crew members working for the involved airlines. Of all the crew members that were invited, a total of 1634 volunteered to participate in the FTL field study; i.e., they registered for participation via an online portal. Four hundred and thirteen (i.e., over 25%) of this group of volunteers provided valuable, adequate data. Internal surveys generally receive a 30 - 40% response rate on average, compared to an average 10 - 15% response rate for external surveys (Duncan, 2008; Fan & Yan, 2010). The invitation to participate in the FTL data collection was sent by the internal airline coordinator, but referred to an external website (of NLR).

[22] We estimated the size of the entire crew population base in D2.2 (Definition of the Data Collection Process). Based on this, we assume to have invited over 47,000 crew members.
[23] See Appendix 3: Training material.

Feedback from the airline coordinators indicated that the number of the 1634 volunteers was perceived to be quite low. They had expected more volunteers from their airlines given the wide range of crew members that were invited to participate. Two, sometimes three reminders were sent by the airline coordinator to encourage crew members to sign up. Also, posters were hung up in the airline crew rooms to promote the FTL study and incentives to participate were offered in the form of a raffle[24] that was run among the 25% of participants within each participating operator airline that uploaded the most valuable data[25].

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.

Measurement protocol

We received many emails and phone calls with questions and concerns about the PVT. Pilots stated numerous times that the PVT was too burdensome and intrusive to perform on the flight deck. Also, as can be determined from the dataset, the timing of when the PVT was performed by the pilots was not (always) in line with the protocol that stated to perform the test 15 minutes prior the TOD of the final sector of that day. In the end, this resulted in inadequate data points for the inclusion of PVT data in the final dataset.

Not only for the PVT, but also for the KSS and SP it turned out to be relatively hard to adhere to the times indicated in the measurement protocol. Raw data showed that crew members could not always complete the KSS and SP at/around 15 minutes prior

[24] A Fitbit activity tracker was used as a prize.
[25] 'Valuable' was the weighted sum of number of days covered, number of assessments and the coverage of sleep/wake logs registering actual sleep.

the TOD. In this context we must add that, from the gathered data it was also hard to determine the exact moment of TOD. We defined this moment as half an hour before wheels on the ground, but this is different for different types of operation. Therefore, we selected the KSS and SP measured closest to this default time, but only when this time was between 90 minutes before wheels on ground until 150 minutes after wheels on ground.

Participants did not always collect data for 14 days. We could see this in the raw data and also from emails received from several participants who indicated they had stopped collecting data early. Reasons that were mentioned are: remaining days were off duty, vacation, or sick leave. There were also cases where participants finished late (after the 14 days) because they included another flight duty.

With regard to the sleep log on the app, some participants had difficulty understanding the difference between adding a sleep/wake log and a sleep period. The protocol asks the user to define in the app the period of sleep and wake during which you wish to keep a record (i.e., the 14-day period of this field study). Within this 14-day period participants had to add their periods of sleep. The sleep/wake log existed to explicitly let the participant state that he/she was awake at times other than the added sleep periods. This was not understood correctly by all participants, based on the raw data and on the questions asked by telephone and email.

Potential events that might undermine the quality metrics of the data collected

In this section, specific emphasis is given to the reporting and appropriate mitigation of any potential events that might undermine the quality metrics of the data collected.

Based on the concerns raised with respect to the PVT measure we decided to not use the PVT as a primary outcome measure in the main analyses. This was in line with what is stated in D2.2 (Definition of the Data Collection Process). The PVT data was however addressed in supplementary analyses in Appendix 2: Additional analyses.

The issue raised with the sleep/wake log had no effect on the quality of the data collected. Participants either logged their sleep using the ‘add sleep’ functionality as intended, or logged their sleep by adding the sleep/wake log periods. It was clear from the raw data that the latter logs were intended as sleep periods. This was supported by the fact that we found was no overlap of these sleep logs with the logged duty periods.

Missing data was coded 9999 to indicate that the data is missing. The complete dataset remained intact for analyses for which the missing data was not essential. If data from essential variables (such as KSS scores at TOD) was missing for a participant, this participant dataset was removed before analysis.

All variables were checked for outliers. A value was considered to be an outlier when the value exceeded three standard deviations from the mean value of this variable. Outliers were re-categorised as missing data.

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.

Åkerstedt, T., Hallvig, D., & Kecklund, G. (2017). Normative data on the diurnal pattern of the Karolinska Sleepiness Scale ratings and its relation to age, sex, work, stress, sleep quality and sickness absence/illness in a large sample of daytime workers. *Journal of Sleep Research, 26(5)*, 559-566.

Belenky, G., Wesensten, N. J., Thorne, D. R., Thomas, M. L., Sing, H. C., Redmond, D. P., Russo, M. B., & Balkin, T. J. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *Journal of Sleep Research, 12(1)*, 1-12.

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

Duncan, D. (2008). The adequacy of response rates to online and paper surveys: what can be done? *Assessment & Evaluation in Higher Education, 33(3)*, 301-314.

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior, 26(2)*, 132-139.

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., Mulrine, H. M., van den Berg, M. J., Smith, A. A., Signal, T. L., Wu, L. J., Belenky, G. (2014). Pilot fatigue: relationships with departure and arrival times, flight duration, and direction. *Aviation Space and Environmental Medicine, 85(8)*, 833-840.

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

Gander, P. H., Mulrine, H. M., van den Berg, M. J., Smith, A. A., 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.

Haavisto, M. L., Porkka-Heiskanen, T., Hublin, C., Härmä, M., Mutanen, P., Müller, K., Virkkala, J., & Sallinen M. (2010). Sleep restriction for the duration of a work week impairs multitasking performance. *Journal of Sleep Research, 19(3)*, 444-454.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*, 3rd edition. Wiley.

Kecklund, G., Sallinen, M., & Axelsson, J. (2017). Optimizing shift scheduling. In M. Kryger, T. Roth (Eds.) Principles and Practice of Sleep Medicine, Sixth edition, Elsevier, pp. 742-749.

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

Powell, D. M., Spencer, M. B., Holland, D., Broadbent, E., & Petry, 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.

Pylkkönen, M., Sihvola, M., Hyvärinen, H. K., Puttonen, S., Hublin, C., & Sallinen, M. (2015). Sleepiness, sleep, and use of sleepiness countermeasures in shift-working long-haul truck drivers. *Accident Analysis & Prevention, 80*, 201-210.

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.

Sallinen, M., Sihvolaa, M., Puttonena, S., Ketolac, K., Tuoric, A., Härmäa, M., Kecklundd, G., & Åkerstedt, T. (2017). Sleep, alertness and alertness management among commercial airline pilots on short-haul and long-haul flights. *Accident Analysis & Prevention*, *98*, 320-329.

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.

Santhi, N., Horowitz, T.S., Duffy, J.F., & Czeisler, C.A. (2007). Acute sleep deprivation and circadian misalignment associated with transition onto the first night of work impairs visual selective attention. *PLoS One, 28, 2(11),* e1233.

Spencer, M. B., & Montgomery, J. M. (1997). Sleep patterns of aircrew on charter/air haulage routes. PLSD Report No. PSLD/CHS5/CR/96/082. Hampshire: Defence Evaluation and Research Agency.

Van den Berg, M. J., Signal T. L., Mulrine, H. M., Smith, A. A. T., Gander, P. H., & Serfontein, W. (2015). Monitoring and managing cabin crew sleep and fatigue during an ultra-long range trip. *Aerospace Medicine and Human Performance*, *86(8)*, 705-713.

Van Dongen, H. P. A., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep, 26*, 117-126.

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.

Zijlstra, F. R. H. (1993). *Efficiency in work behavior. A design approach for modern tools*. PhD thesis, Delft University of Technology. Delft, The Netherlands: Delft University Press.

List of abbreviations

Abbreviations	***Description***
Anova	Analysis of variance
BMI	Body Mass Index
CAT	Commercial Air Transport
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
Max	Maximum
OR	Odds Ratio
ORO	Organisation Requirements (in the air Operations Regulation)
PVT	Psychomotor Vigilance Task
RSME	Rating Scale Mental Effort
RT	Reaction Time
SD	Standard Deviation
SE	Standard Error
SP	Samn-Perelli
TOD	Top Of Descent
WOCL	Window Of Circadian Low

Appendix 1: Variables list

The following list contains all variables that were used in the final dataset. The first column contains abbreviations of the variables; the second column describes the variable names as used in the data analyses software programs; the third column provides a full description of the variables; and the fourth column indicates if the variables was considered as an independent variables for FDP1 (Night duties of more than 10 hours) and/or FDP2 (Disruptive schedules).

Variable name	*Variable name used in analyses*	*Descriptions/Remarks*	*Independent variables FDP1 and FDP2*
General			
Crew_ID	CrewID	Crew member number	
Duty_ID	DutyID	Duty number	
WP_ID	WPNumber	Work period number. A work period is defined as the period between two recovery periods. A recover period is defined as a period of at least 36 hours containing at least two local nights (from 22:00h to 08:00h). Local means the time zone at the location the crew member is during the recovery period	
FDP related variables			
StartH	TimeDayStartAT	Hour at which the FDP starts in the time zone the crew member is acclimatised to	FDP1 & FDP2
EndH	TimeDayEndAT	Hour at which the FDP ends in the time zone the crew member is acclimatised to	FDP1 & FDP2
FDPdur	FDPDuration	Total duty time (in minutes) in FDP	FDP1 & FDP2
FDPdurWk	DutyTime7	Total duty time (in minutes) in 7 preceding days	FDP1 & FDP2
FDPWOCLm	MWOCL	Total time in WOCL (in minutes) in FDP (WOCL is between 02:00h and 05:59h)	FDP1 & FDP2
FDPWOCL%	PWOCL	Percentage of the FDP that takes place during WOCL (between 02:00h and 05:59h). Thus, the following formula was used: MWOCL/FDPDuration	FDP1 & FDP2
FDPWOCL	WOCL01	If at least one minute of the FDP took place in WOCL (between 02:00h and 05:59h)	FDP1 & FDP2
FDPaltWOCLm	FDPMCLow25	Total time in alternative WOCL (in minutes) in FDP (WOCL is now between 02:00h and 04:59h)	FDP1 & FDP2
FDPaltWOCL1h	WOCL50	At least 1 hour of the FDP during alternative WOCL (WOCL is now between 02:00h and 04:59h)	FDP1 & FDP2
FDPsectors	FDPSectors	Total number of sectors flown in current FDP	FDP1 & FDP2
RestPeriod	RestPeriod	Total time off (in minutes) directly prior to FDP that is longer than 10 hours, but shorter than a recovery period	FDP1 & FDP2
TZ_EW TZ_WE	TimeZonesEW TimeZonesWE	Number of time zones crossed from East to West (or vice versa) when comparing start with end of FDP	FDP1 & FDP2
Km_SN Km_NS	DistSN DistNS	Distance (in kilometres) from reference location to most Southern location of FDP or vice versa when comparing start with end of FDP	FDP1 & FDP2
TimeDayTOD	TimeofDayKSSTOD	The time of day (in hours ranging from 0 - 23)	FDP1 & FDP2

		at the time the KSS was rated at TOD. This is the time the crew member is acclimatised to at the time of measurement. If the crew member was unacclimatised the record was removed	
WOCLTOD	WOCLTOD	If the time of day at the time the KSS was rated at TOD fell within the WOCL. This is the time the crew member is acclimatised to at the time of measurement. If the crew member was unacclimatised the record was removed	FDP1 & FDP2
MinPassedTOD	FDPPassed	Duty time passed (in minutes) at TOD	FDP1 & FDP2
FDPmonth	FDPMonth	Month of start of FDP	FDP1 & FDP2

Sleep-wake variables

Sleep24h Sleep48h Sleep72h	HActSleep24 HActSleep48 HActSleep72	Total sleep (in minutes) in the 24, 48 and 72 hours prior to TOD based on sleep log data	FDP1 & FDP2
FDPsleep	FDPNapTime	Total sleep (in minutes) during FDP	FDP1 & FDP2
AwakeTOD	HAwake	Total time (continuous minutes) of wakefulness since the last period of sleep prior to TOD based on sleep log data	FDP1 & FDP2
AwakeStart	HAwakeM	Total time (continuous minutes) of wakefulness since the last period of sleep before start of the FDP based on sleep log data	FDP1 & FDP2

Fatigue measurement variables

KSSMAX	KSSMAX	Karolinska Sleepiness Scale measured from crew member. The scale is a 9-point scale. MAX is defined as the highest KSS score during the FDP. The crew member was instructed to fill out the scale during each flight within the FDP on TOD or, if the flight was long-haul to fill it out during the cruise phase of the flight and at TOD. The highest value that was measured between the start of the duty until 150 minutes after wheels on ground is reported	
KSSTOD	KSSTOD	KSS measured from crew member. The scale is a 9-point scale. TOD is defined at half an hour before wheels on the ground. The KSS measured closest to this time was selected, but only when this time was between 90 minutes before wheels on ground until 150 minutes after wheels on ground. KSSTOD refers to the KSS scored on TOD during the final leg of the FDP	
SPMAX	SPMAX	Samn-Perelli scale measured from crew member. The scale is a 7-point scale. See KSSMAX for explanation on MAX	
SPTOD	SPTOD	SP measured from crew member. See KSSTOD for explanation on TOD	
PVTLap	PVTLap	A lapse during the psychomotor vigilance task is defined as a response longer than 500 milliseconds	
PVTRT	PVTRT	PVT reaction time in milliseconds	
Mental_Effort	ME	The level of mental effort experienced; rated from 0 (not at all effortful) to 150 (tremendously effortful)	FDP1 & FDP2
HassleNo	Hssl	Number of hassle/contributing factors that were applicable during the FDP	FDP1 & FDP2

Grouping variables

FDP1LN	FDP2	At least one minute of the duty takes place between 02:00h and 04:59h and the duration of the duty is at least 10 hours. FDP1 was originally indicated as FDP2 in D1 Addendum; this FDP was ranked first, therefore FDP1 was introduced	
FDP2DS	FDP6	One of the following variables is applies: FDP6ETES, FDP6ETLF, FDP6LTES, FDP6LTLF, FDP6Night. FDP Disruptive Schedules was originally indicated as FDP6 in D1 Addendum; this FDP was ranked second, therefore FDP2 was introduced	
FDP2ETES	FDP6ETES	If the crew member works in a country or the airline's home base is situated in a country that is defined as early type and the FDP started between 05:00h and 05:59h	
FDP2ETLF	FDP6ETLF	If the crew member works in a country or the airline's home base is situated in a country that is defined as early type and the FDP ended between 23:00h and 01:59h	
FDP2LTES	FDP6LTES	If the crew member works in a country or the airline's home base is situated in a country that is defined as late type and the FDP started between 05:00h and 06:59h	
FDP2LTLF	FDP6LTLF	If the crew member works in a country or the airline's home base is situated in a country that is defined as late type and the FDP ended between 00:00h and 01:59h	
FDP2N	FDP6Night	If at least one minute of the FDP took place between 02:00h and 04:59h	
FDP2ES	FDPESFTL	If FDP6ETES or FDP6LTES	
FDP2LF	FDPLFFTL	If FDP6ETLF or FDP6LTLF	
FDPLFN	FDPLFN	FDP late finish plus night	FDP2
FDPES37	FDPES37	FDP early start between 03:00h and 06:59h	FDP2
VeryES	FDPVES	Start FDP between 05:00h and 05:59h	FDP2
ES	FDPES	Start FDP between 06:00h and 06:59h	FDP2
LF	FDPLF	End FDP between 23:00h and 00:29h	FDP2
VeryLF	FDPVLF	End FDP between 00:30h and 01:59h	FDP2
ES2LF	WPESLFFTL	Transition from early start (FDP6ETES or FDP6LTES) to late finish (FDP6ETLF or FDP6LTLF). These are consecutive duties	FDP2
ES2N	WPESNFTL	Transition from early start (FDP6ETES or FDP6LTES) to night (FDP6Night). These are consecutive duties	FDP2
LF2ES	WPLFESFTL	Transition from late finish (FDP6ETLF or FDP6LTLF) to early start (FDP6ETES or FDP6LTES). These are consecutive duties	FDP2
LF2N	WPLFNFTL	Transition from late finish (FDP6ETLF or FDP6LTLF) to night (FDP6Night). These are consecutive duties	FDP2
N2ES	WPNESFTL	Transition from night (FDP6Night) to early start (FDP6ETES or FDP6LTES). These are consecutive duties	FDP2
N2LF	WPNLFFTL	Transition from night (FDP6nights) to late finish (FDP6ETLF or FDP6LTLF). These are consecutive duties	FDP2
ForwardR	WPRotCForw	If WPESLFFTL, WPLFNFTL or WPNESFTL	FDP2
BackwardR	WPRotCCBackw	If WPESNFTL, WPLFESFTL or WPNLFFTL	FDP2

NoConES	CONES	Number of consecutive early starts (FDP6ETES or FDP6LTES). If CONES = 2, this means that the previous FDP started early and that the current FDP started early	FDP2
NoConLF	CONLF	Number of consecutive late finishes (FDP6ETLF or FDP6LTLF). If CONLF = 2, this means that the previous FDP ended late and that the current FDP ended late	FDP2
NoConN	CONN	Number of consecutive nights (FDP6Night). If CONN = 2, this means that the previous FDP was a night duty and that the current FDP is a night duty	FDP1 & FDP2
Individual-related variables			
Function	Function	Pilot or cabin crew	FDP1 & FDP2
Gender	Gender	Male, female, other	FDP1 & FDP2
Age	Age	In years	FDP1 & FDP2
BMI	BMI	Body Mass Index. The BMI is defined as the body mass divided by the square of the body height (kg/m^2)	FDP1 & FDP2
HabSleep	HabSlp	Habitual sleep length in minutes	FDP1 & FDP2
ComTime	ComTime	Commute time (in minutes) from home to work	FDP1 & FDP2
DiurnType	DiurnType	Diurnal type. Diurnal types are grouped as follows: 1. Extreme eveningness 2. Eveningness 3. Intermediate 4. Morningness 5. Extreme morningness	FDP1 & FDP2

Appendix 2: Additional analyses

This section shows additional analyses for the outcome measure KSS, SP, PVT, and contributing factors to workload (or hassle factors).

KSS

Research question

Are KSS scores in FDP1 (Night duties of more than 10 hours) and FDP2 (Disruptive schedules) significantly different from the KSS scores in the FDP Other set?

Participants

The entire crew population as described in the section 'Description of crew member data' was used for this analysis. The FDP Other set consists of all other FDPs than FDP1 and FDP2.

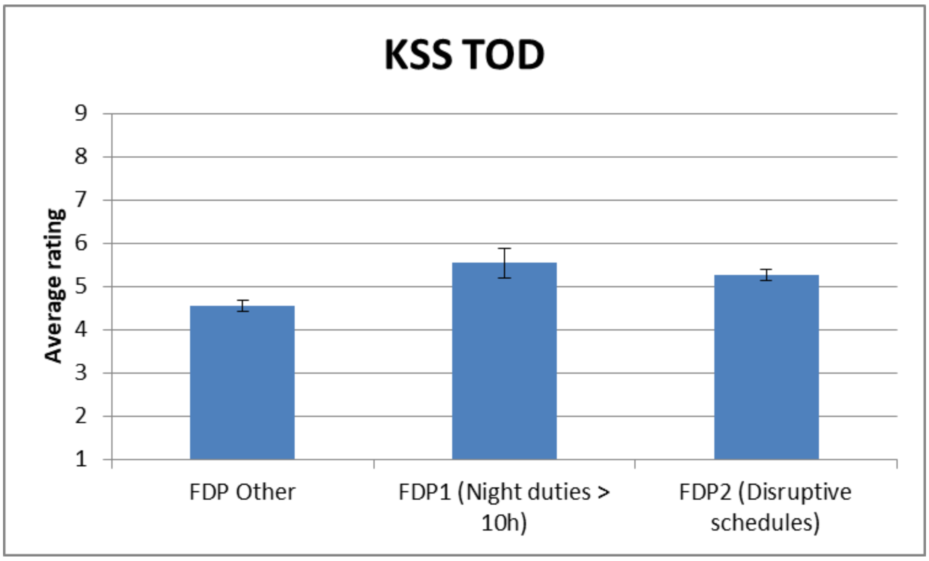
Method

Mean KSS TOD scores for FDP1, FDP2 as well as the FDP Other set were calculated and compared. Anova and non-parametric tests were executed to determine whether KSS scores for both FDPs differed significantly from the FDP Other.

Results

KSS TOD *(1-9)*	***N***	***Minimum***	***Maximum***	***Mean***	***SD***
FDP Other	794	1	9	4.57	1.828
FDP1	146	2	9	5.55	2.041
FDP2	847	1	9	5.27	1.968

SD: Standard Deviation.



Significant differences were found between FDP1 and FDP Other ($F (1, 1639) = 16.595$, $p < .000$), and also between FDP2 and FDP Other ($F (1, 1639) = 55.710$, $p < .000$).

Conclusion

The KSS TOD scores differed significantly between both FDP1 (Night duties of more than 10 hours) and FDP2 (Disruptive schedules) and the FDP Other.

SP

Research question

Are SP scores in FDP1 (Night duties of more than 10 hours) and FDP2 (Disruptive schedules) significantly different from the SP scores in the FDP Other set?

Participants

The entire crew population as described in the section 'Description of crew member data' was used for this analysis. The FDP Other set consists of all other FDPs than FDP1 and FDP2.

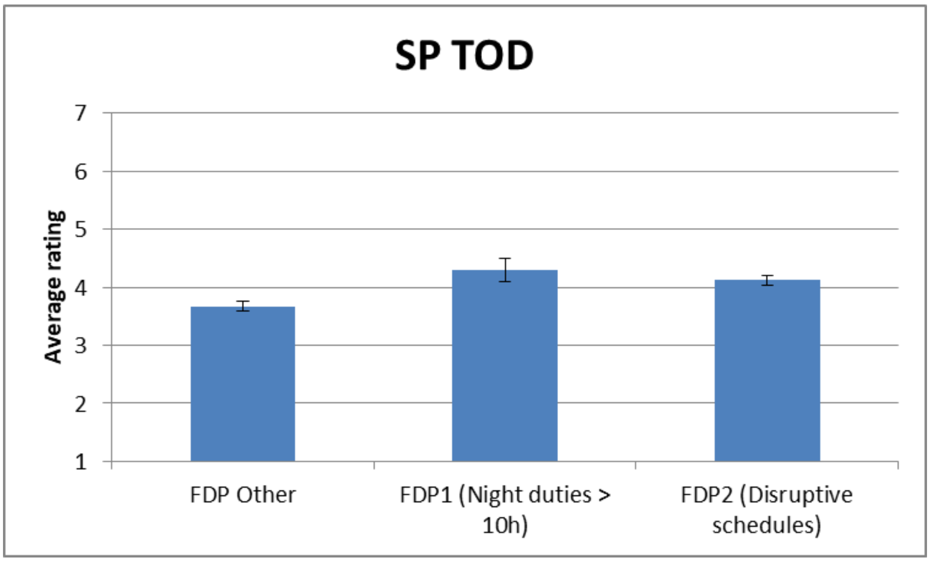
Method

Mean SP TOD scores for FDP1, FDP2 as well as the FDP Other set were calculated and compared. Anova and non-parametric tests were executed to determine whether SP scores for both FDPs differed significantly from the FDP Other.

Results

SP TOD *(1-7)*	***N***	***Minimum***	***Maximum***	***Mean***	***SD***
FDP Other	796	1	7	3.67	1.215
FDP1	144	1	6	4.29	1.256
FDP2	847	1	7	4.13	1.258

SD: Standard Deviation.



Significant differences were found between FDP1 and FDP Other ($F(1, 1641) = 15.074010$, $p < .000$), and also between FDP2 and FDP Other ($F(1, 1641) = 56.185019$, $p < .000$).

Conclusion

The SP TOD scores differed significantly between both FDP1 (Night duties of more than 10 hours) and FDP2 (Disruptive schedules) and the FDP Other.

PVT

Research question

Are PVT scores in FDP1 (Night duties of more than 10 hours) and FDP2 (Disruptive schedules) significantly different from the PVT scores in the FDP Other set?

Participants

The entire crew population as described in the section 'Description of crew member data' was used for this analysis. The FDP Other set consists of all other FDPs than FDP1 and FDP2.

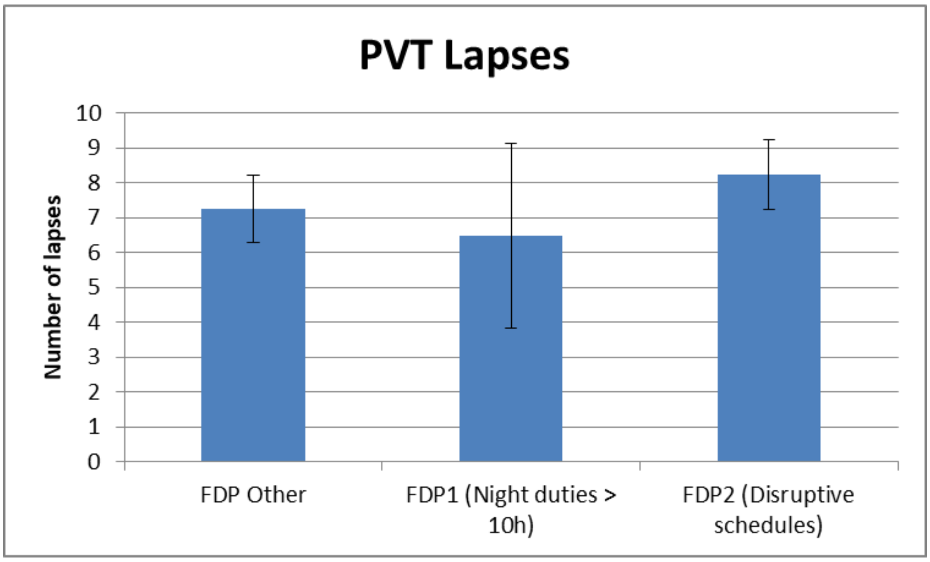
Method

Mean PVT scores (number of Lapses and Reaction Time) for FDP1, FDP2 as well as the FDP Other set were calculated and compared. Anova and non-parametric tests were executed to determine whether PVT scores for both FDPs differed significantly from the FDP Other.

Results

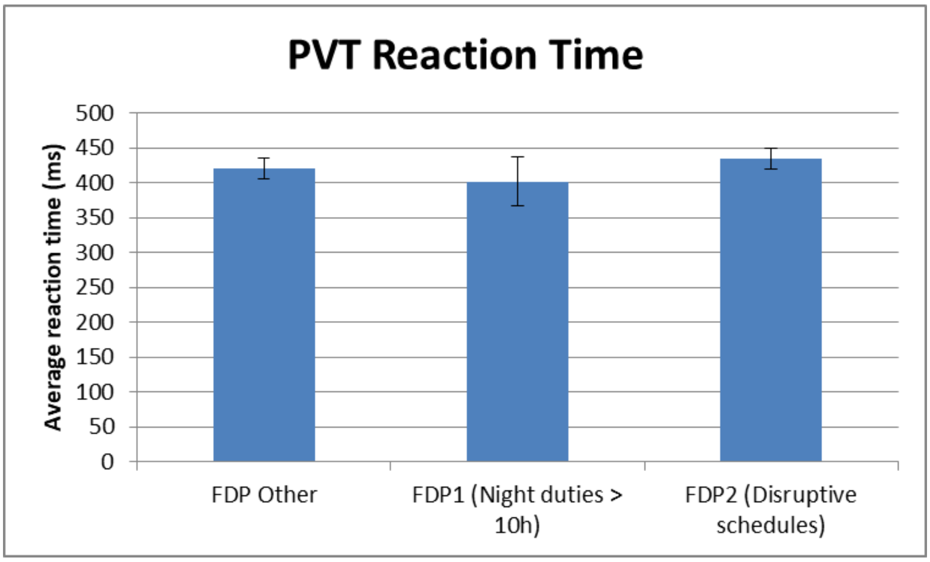
PVT Lapses	***N***	***Minimum*** *No.*	***Maximum*** *No.*	***Mean*** *No.*	***SD***
FDP Other	440	0	45	7.24	10.296
FDP1	71	0	45	6.48	11.442
FDP2	465	0	45	8.24	10.860

SD: Standard Deviation.



PVT RT	***N***	***Minimum*** *Milliseconds*	***Maximum*** *Milliseconds*	***Mean*** *Milliseconds*	***SD***
FDP Other	438	262	1262	419.90	158.164
FDP1	68	276	874	401.93	149.520
FDP2	460	0	1388	434.60	161.708

SD: Standard Deviation.



No significant differences were found between the conditions for both PVT Lapses and Reaction Time.

Conclusion

The PVT scores (number of Lapses and Reaction Time) did not differ significantly between both FDP1 (Night duties of more than 10 hours) and FDP2 (Disruptive schedules) and the FDP Other.

Contributing factors

Research question

What are the contributing factors to workload (or hassle factors) within FDP1 (Night duties of more than 10 hours) and FDP2 (Disruptive schedules)?

Participants

The entire crew population as described in the section 'Description of crew member data' was used for this analysis. The FDP Other set consists of all other FDPs than FDP1 and FDP2.

Method

Mental effort was considered in the field data collection by means of the RSME; that is, the level of Mental effort experienced by the participant in the particular duty period was rated by moving a slide bar. The bar ran from 0 (not at all effortful) to 150 (tremendously effortful).

In case participants scored 50 (somewhat effortful) or higher, a list with different contributing factors popped up. Participants were invited to select the factors that contributed in any way to their workload on that particular duty period. The full list of contributing factors is defined and presented in D2.2 (Definition of the Data Collection Process).

Results

In total 521 FDPs were found with Mental effort scores of 50 or higher. Contributing factors were scored for 457 FDPs of these 521 FDPs. Participants indicated in total 1697 factors.

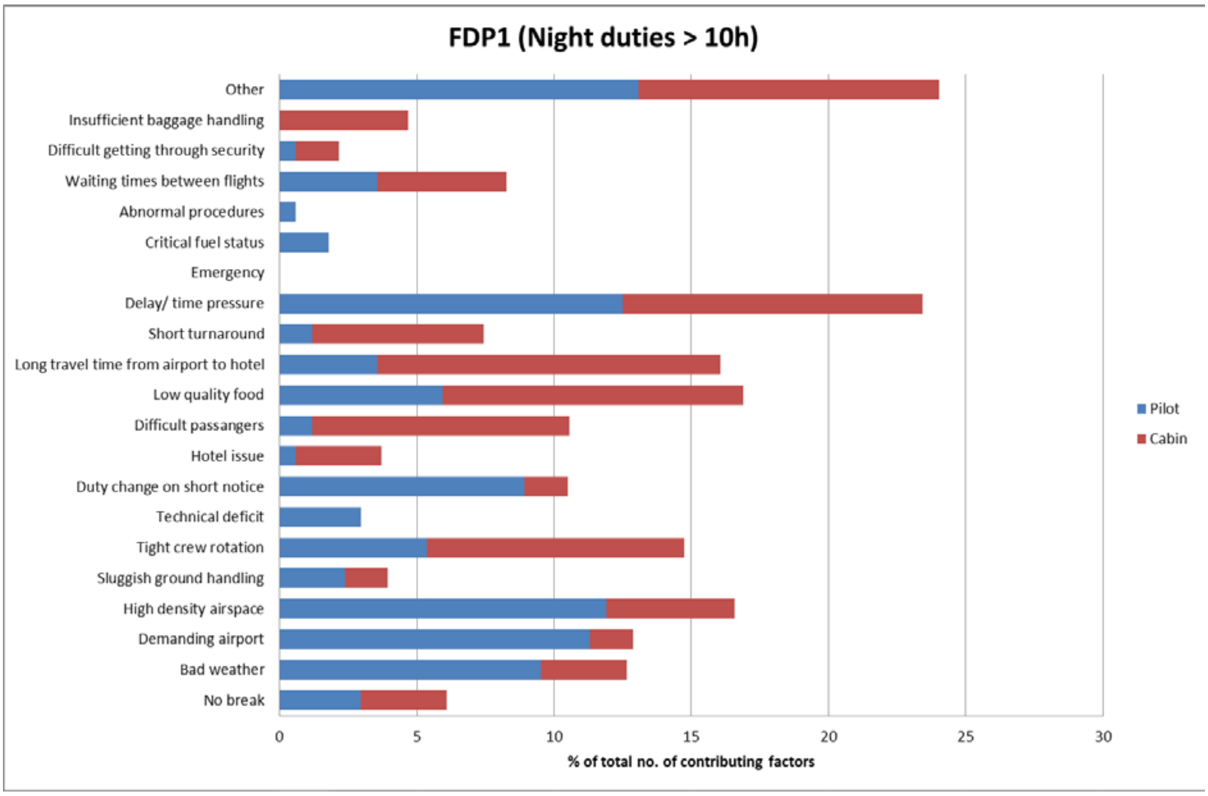


Figure 26 Contributing factors to workload for FDP1 (Night duties of more than 10 hours)

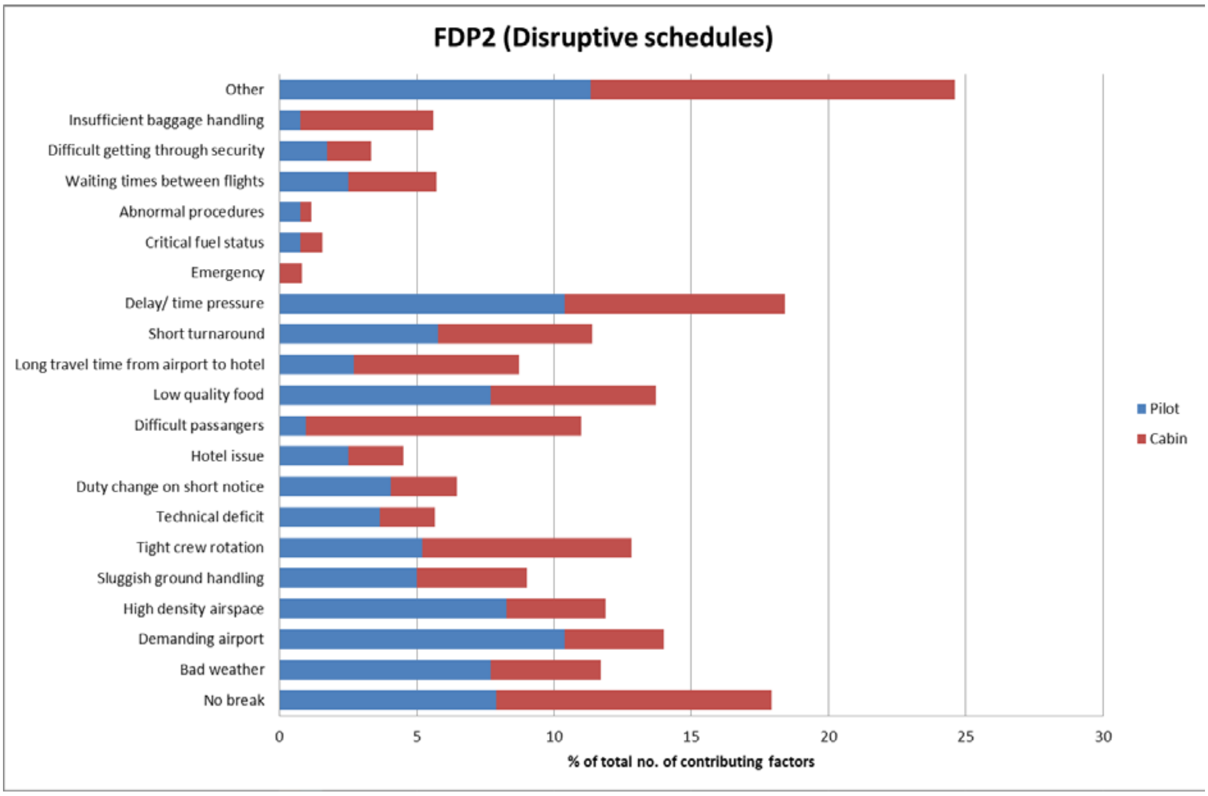


Figure 27 Contributing factors to workload for FDP2 (Disruptive schedules)

Figure 26 and Figure 27 respectively show for FDP1 and FDP2 the percentage of the total number of contributing factors that were indicated by the pilots and cabin crew in the field data collection.

Conclusion

The results for FDP1 (Night duties of more than 10 hours) and FDP2 (Disruptive schedules) on the number of contributing factors to the experienced workload showed a similar trend in a sense that the following factors were scored more (above 10%) than the rest (below 10%):

- Other;
- Delay/time pressure;
- Low quality food;
- Difficult passenger;
- Tight crew rotation;
- High density airspace;
- Demanding airport; and
- Bad weather.

The factors 'short turnaround' and 'no break' showed relatively higher scores for FDP2 as compared to FDP1; the factors 'long travel time from airport to hotel' and 'duty change on short notice' showed relatively higher scores for FDP1 as compared to FDP2.

Appendix 3: Training material

Dedicated to innovation in aerospace

EC/EASA - Air Crew Data Collection Campaign

FTL training module

Training module data collection campaign

Overview FTL Data Collection Campaign

- ✓ Voluntary participation
- ✓ Data collection via the CrewAlert Lite app (iOS) and Actiwatch
- ✓ **Survey code** required in the CrewAlert Lite app
- ✓ Total of **14 consecutive days**, starting with a period of 2 days off
    - Enter and maintain your flight duties and sleep/wake history
    - If invited - Wear an Actiwatch (Note: not all participants will be asked to wear an Actiwatch)
    - Score your alertness (KSS & SP) daily at duty start, cruise (only long-haul), and 15 min prior to top of descent on each flight
    - For pilots only - Test your reaction time (NPI) at duty start and 15 min prior to top of descent on the final flight
- ✓ Don't give in if you miss out on a few data points – the data is still useful
- ✓ **Upload your CrewAlert data** at the end of the collection period

Background to the EU FTL Research Study

The European Committee (EC) together with EASA has set up a research study to perform **a review of the effectiveness of the flight and duty time limitations and rest requirements** applicable as of 18 February 2016.

The objective is to **collect aircrew data on fatigue, alertness, workload and sleep** to determine whether these rules provide sufficient protection from potential consequences of aircrew fatigue. If necessary, recommendations will be drafted for changes to the rules.

This review of the Flight Time Limitations (FTL) is being performed by a study team with the Netherlands Aerospace Centre NLR, in collaboration with Stockholm University, German Aerospace Centre DLR, and Jeppesen.
The EC is paying for the costs of this study

Data Collection Campaign

- ✓ The FTL data collection campaign asks your **participation for 14 days**
- ✓ Each day you are asked – during your normal flight duties and during days off – to regularly fill in two different rating scales, to continuously wear an Actiwatch (if provided to you), to maintain your flight duties and sleep/wake history, and to perform a reaction time test (if cockpit crew)
- ✓ In total this takes about **15 minutes of your time per day**
- ✓ The assessments are performed on the CrewAlert Lite application. For this you need the availability of an **iPhone, iPad or iPod Touch**
- ✓ The data collection is never to conflict with the performance of your duties

Usage and Storage of your Data

- ✓ The gathered data will remain confidential and anonymous since your name is irrelevant to the needs of the research study
- ✓ At the end of the 14 day study period you need to upload your data
- ✓ The data will be transferred (via encrypted technology) over a Wi-Fi connection and stored in a password-protected central database that can be accessed by study team members only
- ✓ The collected datasets will not be disclosed to third parties during or after the study's lifetime or beyond

Setting up the CrewAlert Lite App

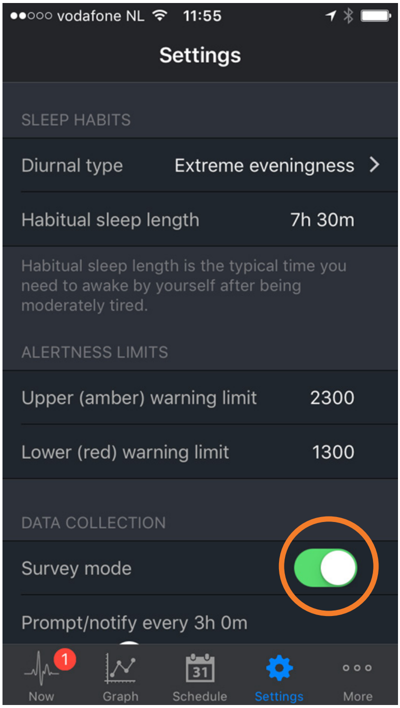
Installing CrewAlert

- ✓ Make sure that you have Internet access
- ✓ Install CrewAlert Lite from the iTunes Appstore
    - For finding CrewAlert Lite on an iPad; please toggle for searching for iPhone apps
    - If you are already using CrewAlert Lite, update your version
    - No need to use the Pro version
- ✓ Please follow the instructions on the required settings on the following pages

CrewAlert settings

Go to “Settings” at the bottom of your screen

- ✓ Set your “Default time base” and “Home base time zone”
- ✓ Fill in your “Transfer Times”
- ✓ Fill in your “Sleep Habits”
- ✓ Toggle “Data Collection” in “Survey mode”
- ✓ Complete your “Demographic Information”



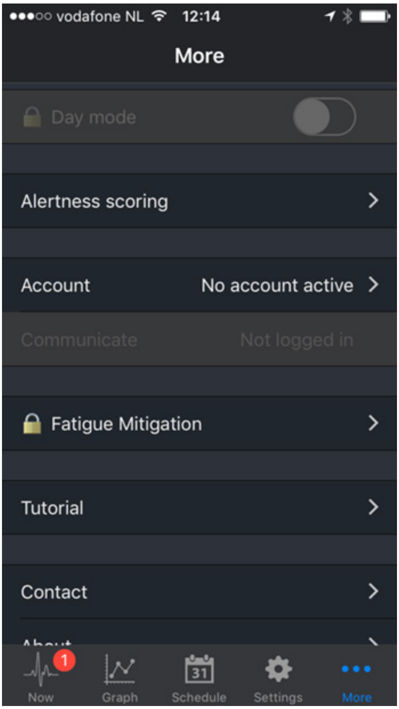
SETTINGS ONLY NEED TO BE FILLED IN ONCE AT THE START OF THE DATA COLLECTION

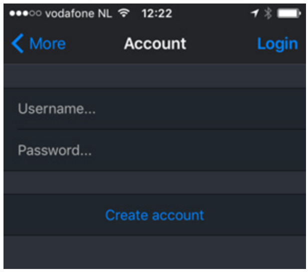
Creating an account

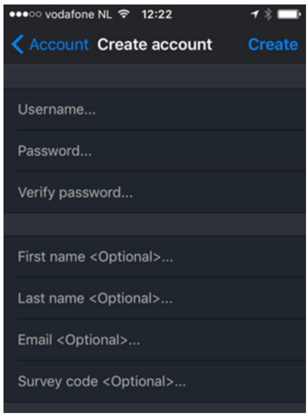
Go to “More” at the bottom of your screen

- ✓ Create an “Account”
- ✓ Fill in *any* username and password
    - No need to fill in “First name”, “Last name” or “Email”
    - Note that the password is NOT the same as your “Survey code”
    - Do not use characters in CAPS for the password
- ✓ Insert the “Survey code” that you received
    - If you have not received a survey code after signing up, please contact the principal investigator via henk.van.dijk@nlr.nl

THE SURVEY CODE IS REQUIRED TO FILTER DOWN THE DATASET TO THE FTL STUDY







Actiwatch operating instructions

How to handle the Actiwatch

NOT ALL PARTICIPANTS WILL BE ASKED TO WEAR THE ACTIWATCH

✓ Wear it for the full 14-day period
✓ Wear it on the wrist of your preference
✓ Press the button each time you plan to sleep or take a nap

✓ The watch is water resistant but … take it off when you take a shower
✓ It records data on an internal chip, it does not transmit data
✓ After the 14 days, hand it in to your FTL airline coordinator or send it to NLR

Daily data collection

Start with 2-days off duty!

✓ Day 1 is used for familiarization with the CrewAlert app and the different measures; day 2 is to gather baseline measures

✓ During off-duty, besides keeping a sleep/wake log and (if invited) wearing the Actiwatch, you are asked three times a day (morning, afternoon, evening) to:
- Rate your level of sleepiness and fatigue on the two rating scales (KSS & SP)
- If cockpit crew - Measure your alertness with the 5-min response time test (NPI)

NOTE: THE NPI TEST IS ONLY REQUIRED IF YOU ARE A PILOT

During each of the 14 days you will be asked to do the following:

After you wake up

- ✓ Fill in the sleep/wake log on the app
- ✓ Rate your level of sleepiness and fatigue on the two rating scales (KSS & SP)
- ✓ *Pilots only:* Measure your alertness with the 5-min reaction time test (NPI)

During your flight duty

- ✓ Rate the KSS & SP in cruise (only long-haul) and 15 min prior the top of descent of each flight
- ✓ *Pilots only:* Perform the NPI 15 min prior the top of descent of your final flight of that day

Planning a nap

- ✓ *If invited:* Press the button on the Actiwatch, indicating that you are going to sleep
- ✓ Fill in the sleep/wake log, after taking a nap

After your flight duty

- ✓ Fill in your achieved flight duty of that day

When going to sleep

- ✓ Fill in the sleep/wake log
- ✓ *If invited:* Press the button on the Actiwatch

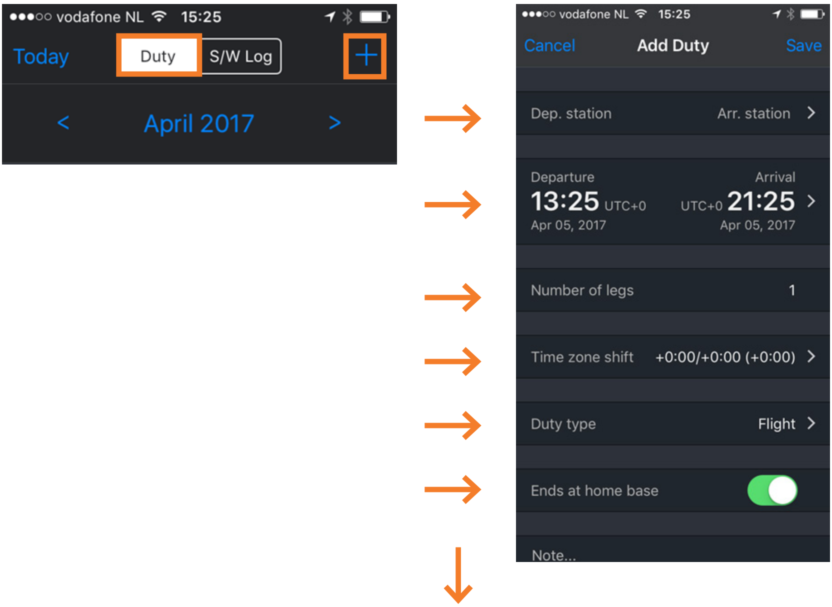
1. Fill in Duty - CrewAlert

Go to “Schedule” at the bottom of your screen

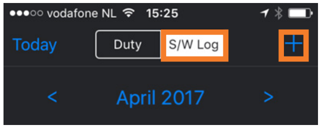
✓ Toggle to “Duty”

✓ Click on “+” to add your latest duty

- Fill in your “Dep. & Arr. station”
- Fill in time start (“Departure”) and end (”Arrival”) of your duty
- Fill in the “Number of legs” within this duty
- Change “Time zone shift” if relevant
- Change “Duty type” if relevant
- Toggle to green if your duty “Ends at home base”
- Fill in “Note...” in case you filed in sick/not fit-to-fly
- Update “No in-flight sleep” if relevant
- Rate your mental effort (“RSME”) within this duty
- Toggle the “Contributing Factors” if relevant (appears after certain level of RSME Mental Effort)



2. Fill in Sleep/Wake Log - CrewAlert



Go to “Schedule” at the bottom of your screen

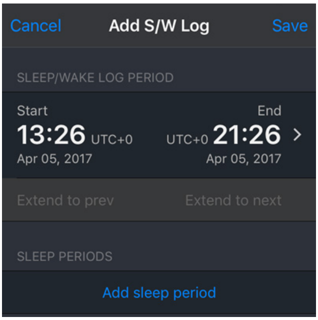
✓ Toggle to “S/W Log”

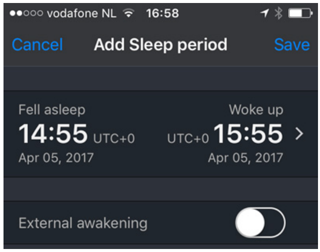
✓ Click “+” to add your latest sleep period

- The S/W Log start and end time define the period of sleep and wake during which you wish to keep a record (i.e. the 14-day period of this campaign)

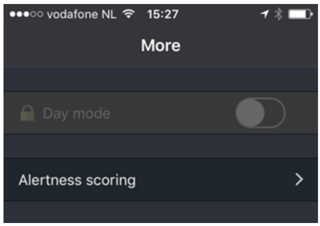
✓ “Add sleep period” within your S/W Log

- Fill in time you “Fell asleep”
- Fill in time you “Woke up”





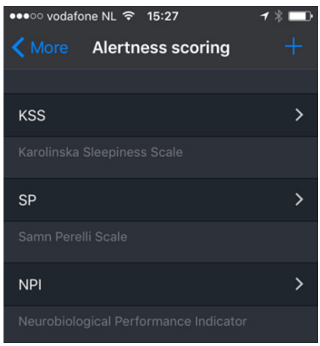
3. Rate your Alertness - CrewAlert



Go to “More” at the bottom of your screen

✓ Click “Alertness scoring”

- Rate the “KSS”
- Rate the “SP”
- Execute the “NPI”: **click “start” and react to the stimulus as quickly as possible for 5 min by touching the black dot** *(only required for pilots)*



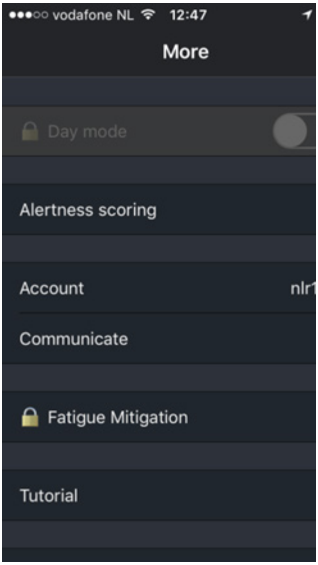
Data upload

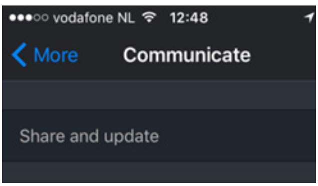
Uploading your CrewAlert Data

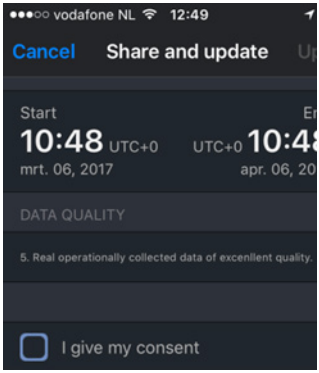
Go to “More” at the bottom of your screen

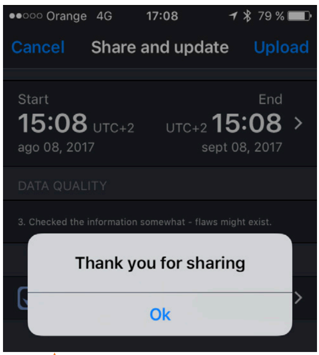
- ✓ Click “Communicate”
- ✓ Click “Share and update”
- ✓ Select from “Data Quality” the fitting description of your data
- ✓ Give your consent to share your data with the study team
- ✓ Your data will not be used if you do not provide consent for this

PLEASE SHARE YOUR DATA AT THE END OF THE 14-DAYS PERIOD









This is the confirmation notice that you will get after uploading your data correctly

Actiwatch Data

- ✓ The data is stored on the Actiwatch itself, uploading is not necessary
- ✓ The number of available Actiwatches is limited due to budgetary constraints

PLEASE RETURN THE ACTIWATCH AS SOON AS POSSIBLE AT THE END OF THE 14-DAYS PERIOD TO YOUR FTL CONTACT OR TO NLR

COMMON MISTAKES:

- ✓ **Thinking a sleep/wake log equals sleep.**
  A sleep/wake log explains sleep and wake to CrewAlert and contains only wake as default. Sleep periods need to be added inside the log to record your actual sleep. (Logged sleep periods are blue in the Graph tab)
- ✓ **Not confirming in-flight sleep with a sleep/wake log.**
- ✓ **Feeding in the wrong date or time reference.**
  Normally you will spot errors in data by a duty turning red. Please investigate and try to correct duties and sleep.
- ✓ **Entering every flight.**
  It is enough to enter each duty: first departure, last arrival, number of flights. This is much quicker!
- ✓ **Becoming tired of collecting data.**
  Take a break in the collection during days off or reduce workload by feeding in less (like skipping NPI). Duties, sleep, SP, and KSS have the highest priority.

Dedicated to innovation in aerospace

Use the return envelope provided by you at distribution.

Please return to:

Netherlands Aerospace Center NLR
Attn: Dr Henk van Dijk
P.O. Box 90502
1006 BM Amsterdam
The Netherlands

THANK YOU FOR PARTICIPATING!

If you have any questions, please contact the principal investigator
Dr Henk van Dijk
NLR - NETHERLANDS AEROSPACE CENTRE
henk.van.dijk@nlr.nl / +31885113504

European Commission