# Automated Collection of Fatigue Ratings at the Top of Descent: A Practical Commercial Airline Tool

David M. C. Powell, Mick B. Spencer, and Keith J. Petrie

Powell DMC, Spencer MB, Petrie KJ. *Automated collection of fatigue ratings at the top of descent: a practical commercial airline tool.* **Aviat Space Environ Med 2011; 82:1–5.**

***Introduction:*** There is a need to develop an efficient and accurate way of assessing pilot fatigue in commercial airline operations. We investigated the validity of an automated system to collect pilot ratings of alertness at the top of descent, comparing the data obtained with existing results from previous studies and those predicted by the validated SAFE fatigue model. ***Methods:*** Boeing 777 pilots were prompted to enter a Samn-Perelli fatigue scale rating directly into the flight management system of the aircraft shortly prior to descent on a variety of short- and long-haul commercial flights. These data were examined to evaluate whether the patterns were in line with predicted effects of duty length, crew number, and circadian factors. We also compared the results with data from previous studies as well as SAFE model predictions for equivalent routes. ***Results:*** The effects of duty length, time of day, and crew complement were in line with expected trends and with data from previous studies; the correlation with predictions from the SAFE model was high (r = 0.88). Fatigue ratings were greater on longer trips (except where mitigated by adding an extra pilot) and on overnight sectors (4.68 vs 3.77). ***Discussion:*** The results suggest that the automated collection of subjective ratings is a valid way to collect data on fatigue in an airline setting. This method has potential benefits for the crew in assessing fatigue risk prior to approach, as part of a fatigue risk management system, with the possibility of wider safety benefits.

**Keywords:** fatigue, intervention, work hours, circadian rhythm, duty time limitations.

An IMPORTANT ISSUE in commercial airline operations is the evaluation of the effect of different sectors and work patterns on pilot fatigue. This information is useful for identifying problem sectors and duties that may compromise the safety of the operation. Such data would also be useful for monitoring changes in fatigue that may indicate the need for an intervention, such as adjusting departure times, re-specifying preflight rest provisions, or perhaps adding an additional pilot (4,5).

One of the difficulties of collecting ongoing fatigue measures from pilots is that the methodologies currently available are labor-intensive and costly. Typically, these have involved researchers' either accompanying crew on a duty and collecting fatigue ratings over the course of each sector, or briefing crew prior to departure to fill out various fatigue ratings and reaction time tasks during each flight (8). While these approaches produce valid estimates of fatigue levels on specific duties, they are impractical to use across the whole of a large airline's operation. What is needed to assess fatigue across the whole of a commercial airline's operation is a valid measure of fatigue at critical phases of flight that can be completed by crew quickly and easily without the use of other personnel or equipment.

In this study we evaluated the validity of pilots' entering their fatigue rating directly into the flight management system of the aircraft just prior to the top of descent. To do this we designed a system that prompts pilots to enter a Samn-Perelli fatigue rating scale (10) for each crewmember into a special screen. This allowed for the routine collection of fatigue data from the crews of those aircraft fitted with this modification.

We evaluated the validity of this methodology in a mixture of short- and long-haul operations. Firstly, we predicted that fatigue levels in the automated top-of-descent ratings would be greater in long-haul flights (operating across time zones and usually at night) than in short-haul duties (generally daylight hours, returning to home base). Further, we predicted that in short-haul flights, fatigue levels would be greater at the end of duty, on the return sector, compared to the end of the outward sector. We also predicted that in long-haul flights, fatigue would be greater for the daylight flights than for the otherwise similar overnight flights and, finally, that the rating method would reflect the predicted beneficial effect of a fourth pilot on longer duties. We compared the results collected by the automated top-of-descent method with data gathered in previous studies using questionnaires and with predictions made by the widely used fatigue predictive model, SAFE (1).

## METHODS

### *Subjects*

Pilots flying Air New Zealand 777-200 aircraft were contacted in advance of the study and invited to participate. All participants were pilots already scheduled to fly the studied patterns on the days available for testing;

From Air New Zealand, Auckland, New Zealand; the University of Otago, Wellington, New Zealand; and the University of Auckland, Auckland, New Zealand.

This manuscript was received for review in May 2011. It was accepted for publication in August 2011.

Address correspondence and reprint requests to: David M. C. Powell, Aviation Medicine Specialist, Air New Zealand, Private Bag 92007, Auckland 1142, New Zealand; david.powell@airnz.co.nz.



DOI: 10.3357/ASEM.3115.2011

no pilots were excluded from testing. Data were collected noninvasively and anonymously during the normal work of the pilots, and participation was voluntary. No demographic data were collected from the subjects and, as the collection of data was anonymous, it is not known which pilots chose to participate.

*Procedure*

The modification to enable the top-of-descent fatigue ratings was made on the flight management system software of Boeing 777-200 aircraft. This aircraft was operating a mixture of long-haul (10–13 h flight time) international sectors and out-and-back duties from home base in New Zealand to destinations in Australia and the Pacific Islands, returning to home base within the same duty period. The long-haul international sectors, with only one exception, operated through the night, with three or four pilot crews to allow each pilot the opportunity to sleep in the aircraft crew rest compartment. The exception was a daylight flight from New Zealand to Japan, with three pilots. The short-haul duties were out-and-back duties, conducted between the hours of 0750 (earliest scheduled departure) and 2000 (latest scheduled arrival). Some of these were flown by two pilots, without crew rest; however, the flights between Auckland and the Cook Islands, and on occasions those between Auckland and Melbourne, being slightly longer, had three pilots, allowing each pilot a rest period away from the flight deck.

*Measures:* An input page was programmed into the aircraft flight management system of the aircraft, asking pilots to input their alertness levels according to the 7-point Samn-Perelli fatigue scale (10). The input screen is shown in **Fig. 1**. The pilots were prompted, at 20 min prior to descent, to conduct the procedure. This involved

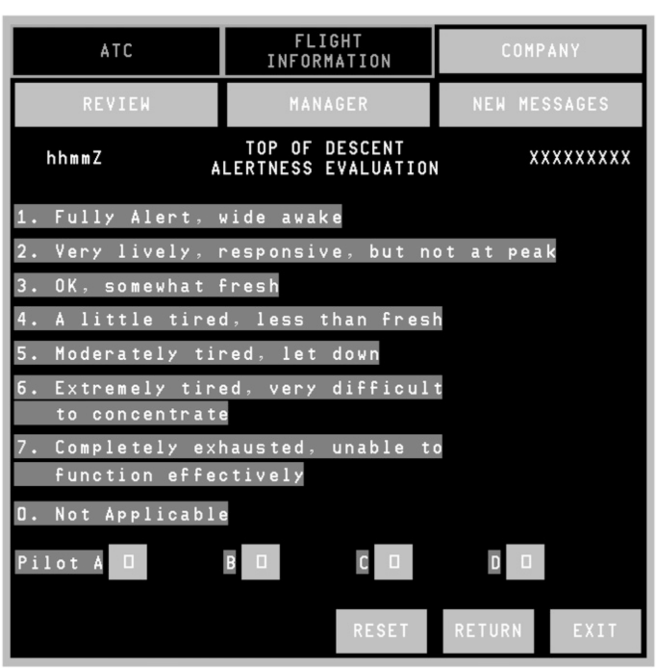


**Fig. 1.** Input screen.

the pilots' discussing their alertness ratings and one of them entering the scores on the appropriate screen.

For each pilot a score was entered anonymously into one of the available boxes. To avoid the effects of sleep inertia, pilots were asked not to provide a rating within 15 min after waking from bunk rest. Once entered, the data were transmitted to a ground-based station, then de-identified by removal of the fields containing date (day of the month) and aircraft registration; the information added to the database consisted of the airports of departure and destination, airline flight number, time of day, and month/year, along with the alertness scores.

*Statistical Analysis*

The returns from individual flights were combined into groups corresponding to the different routes. Then an analysis of variance procedure was used to split the sum of squares due to the means over individual flights (weighted by the number of scores on each flight), into two separate sums of squares, one between routes, the other within routes. Mean values over routes were compared using an error term obtained from the flights-within-routes sum of squares.

Planned comparisons were made as follows: between all long-haul and short-haul routes; between mean scores for the four short-haul routes (eight flights, four out and four back); between the outward long-haul flight to Tokyo and the outward long-haul flight to Hong Kong, which were of similar duration but at different times of day; and, finally, between the long-haul four-pilot flights to Vancouver and Beijing, which departed in the evening with a four-pilot crew, and the (slightly shorter) flight to Hong Kong, which departed in the late evening with a three-pilot crew.

A comparison with existing data was based on the results from two-pilot operations. In a previous study (7), information on fatigue, collected from pilots at top of descent using standard fatigue questionnaires, was summarized in the form of a series of trend curves which related fatigue to the timing and duration of the duty period. The output from this representation, extrapolated where required for longer flights, was compared with the fatigue ratings for all the outbound flights in this study. Inbound flights were excluded due to potential confounding effects of time-zone change and/or multiple sectors (6).

Finally, a comparison was made between the automated top-of-descent ratings for the 23 routes and the predictions of the SAFE fatigue model (2), which has been validated with respect to air transport operations. As some of the parameters used by the model varied within many of the routes (e.g., layover duration, crew size, relief or main crew), the predictions were based on average values using estimates for the distribution of the parameters within the sample.

## RESULTS

Over a period of 1 yr, 4629 ratings were obtained; this represents well over 50% of available flights and a response rate of approximately 38% of the pilots who could have participated. A summary of the fatigue ratings for the individual routes collected with the automatic process is presented in **Table I**. The means and variances for the 23 routes is given in **Fig. 2**, where the routes have been ordered with respect to their mean rating, from the least fatiguing (Auckland–Fiji) to the most fatiguing (Hong Kong–Auckland).

There were clear differences relating to the type of flight undertaken. The highest set of scores was obtained at the end of long-haul nighttime sectors. The lowest scores were the first (outbound) sector of short-haul daylight trips, followed by the second (return) sector. An intermediate rating was obtained from the sole long-haul daytime sector of Auckland-Tokyo. Overall, fatigue levels on the long-haul routes were significantly higher than on the short-haul routes [$F(1,1491) = 1850.7$, $P < 0.001$].

We analyzed the short-haul flights more closely (Fig. 2B) to determine whether the onboard fatigue assessment reflected the expected variations with duty pattern. The scores on return were considerably higher than at the end of the outward flight [$F(1,1491) = 209.8$, $P < 0.001$]. There were also significant differences between the four individual routes [$F(3,1491) = 11.9$, $P < 0.001$] on both the outward and return flights: the Brisbane flights were less fatiguing than those to and from Melbourne ($P < 0.01$), which departed the earliest, and the Fiji flights, which departed latest (at noon) were less fatiguing than those involving the other three short-haul destinations (Brisbane $P < 0.05$; Cook Islands and Melbourne $P < 0.01$). The Cook Islands flights and, on some occasions the Melbourne flights, had three pilots, whereas the other short-haul flights had two; however, since the flights were during daylight, the presence of the third pilot was unlikely to have resulted in bunk sleep.

To examine whether the automated alertness assessment was sensitive to the effects of time of day, we compared an overnight flight with a daytime flight of similar duration. The average fatigue score at the end of the overnight Auckland to Hong Kong flight was significantly higher than at the end of the daytime Auckland to Tokyo flight, which was of similar duration [4.68 vs. 3.77; $F(1,1491) = 96.3$, $P < 0.001$].

We also tested whether the automated alertness assessment showed the effect of an additional pilot by comparing evening flights from Auckland-Vancouver and Auckland-Beijing with four-pilot crews with an evening three-pilot Auckland-Hong Kong flight. Fatigue using this assessment method was significantly higher at the end of the flight to Hong Kong with a three-pilot crew [4.21 vs. 4.68; $F(1,1491) = 16.4$, $P < 0.001$] than at the end of the four-pilot flight to Vancouver; however, this flight departed earlier in the evening. The Beijing flight was a later departure, like Hong Kong and, although it carried an extra pilot, there was no significant difference between these two flights.

We previously examined Samn-Perelli ratings completed by pilots on paper at the top of descent from regional two-pilot operations. From these results we derived a set of trend curves based on start time and approximate duty duration (7). In this study we compared those trend curves to the fatigue scores obtained by the automated collection method. The comparison between the fatigue scores on the outward flights with those from the predictions derived from previous data is illustrated in **Fig. 3**. The two-crew flights were generally in very close agreement. However, the fatigue scores on the three and four pilot overnight flights, which allow for

**TABLE I. RESULTS OF FATIGUE RATINGS FOR INDIVIDUAL ROUTES.**

| From | To | Takeoff (Approx. Local Time) | Flight Duration (h) | No. of Pilots | Type of Flight | Prior Nights Layover | No. of Flights | Mean | Variance |
|---|---|---|---|---|---|---|---|---|---|
| Auckland | Fiji | 1200 | 3.0 | 2 | S/H out | 0 | 36 | 1.88 | 0.69 |
| Auckland | Brisbane | 1000 | 3.5 | 2 | S/H out | 0 | 69 | 2.18 | 0.94 |
| Auckland | Cook Islands | 1100 | 3.8 | 3 | S/H out | 0 | 34 | 2.31 | 0.82 |
| Auckland | Melbourne | 0800 | 3.8 | 2/3 | S/H out | 0 | 75 | 2.56 | 1.12 |
| Fiji | Auckland | 1600 | 3.0 | 2 | S/H back | 0 | 37 | 2.89 | 0.61 |
| Brisbane | Auckland | 1300 | 3.0 | 2 | S/H back | 0 | 75 | 3.20 | 0.99 |
| Melbourne | Auckland | 1200 | 3.4 | 2/3 | S/H back | 0 | 94 | 3.43 | 0.69 |
| Cook Islands | Auckland | 1700 | 3.5 | 3 | S/H back | 0 | 37 | 3.45 | 0.68 |
| Auckland | Tokyo | 1000 | 11.2 | 3 | L/H out | 0 | 104 | 3.77 | 0.69 |
| Auckland | Vancouver | 2000 | 13.2 | 4 | L/H out | 0 | 26 | 4.21 | 0.49 |
| Auckland | Shanghai | 0000 | 12.5 | 3/4 | L/H out | 0 | 63 | 4.36 | 0.61 |
| Auckland | San Francisco | 2000 | 12.2 | 3/4 | L/H out | 0 | 92 | 4.44 | 0.62 |
| Auckland | Beijing | 2300 | 13.5 | 4 | L/H out | 0 | 34 | 4.55 | 0.52 |
| Auckland | Hong Kong | 0000 | 11.5 | 3 | L/H out | 0 | 129 | 4.72 | 0.64 |
| Hong Kong | London | 0800 | 13.2 | 3 | L/H out | 1,2+ | 99 | 4.40 | 0.47 |
| Vancouver | Auckland | 2000 | 14.0 | 4 | L/H back | 2+ | 21 | 4.44 | 0.74 |
| Tokyo | Christchurch | 1800 | 12.0 | 3 | L/H back | 1,2+ | 33 | 4.52 | 0.83 |
| Beijing | Auckland | 1200 | 13.2 | 4 | L/H back | 2+ | 40 | 4.62 | 0.71 |
| Tokyo | Auckland | 1800 | 11.0 | 3 | L/H back | 1,2+ | 83 | 4.62 | 0.45 |
| Shanghai | Auckland | 1400 | 11.5 | 3/4 | L/H back | 2+ | 57 | 4.65 | 0.40 |
| San Francisco | Auckland | 2000 | 13.2 | 3/4 | L/H back | 2+ | 106 | 4.68 | 0.50 |
| London | Hong Kong | 2100 | 12.2 | 3 | L/H back | 2+ | 91 | 4.73 | 0.60 |
| Hong Kong | Auckland | 1800 | 10.8 | 3 | L/H back | 1,2+ | 148 | 4.81 | 0.56 |

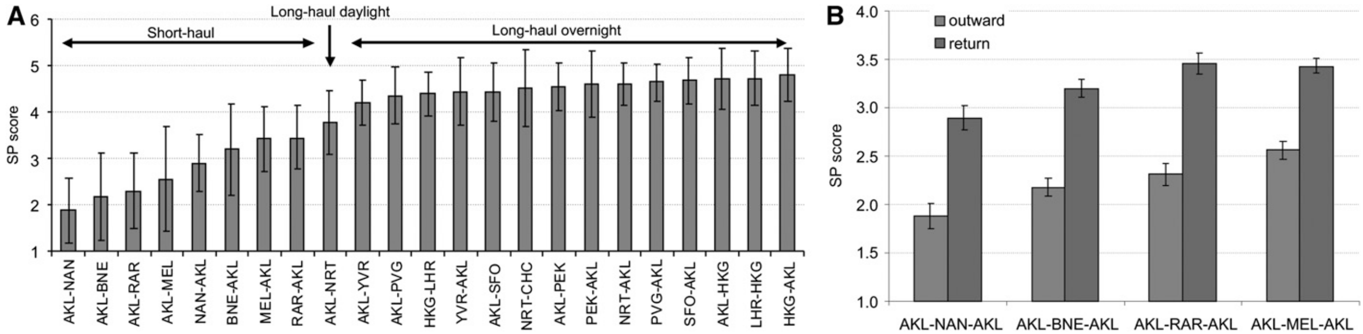


**Fig. 2.** Samn-Perelli fatigue scores by sector (mean ± SE). A) All duties; B) out-and-back daylight duties.

bunk rest, were consistently lower than predictions which were based on a two-crew operation. We also compared the average scores for the 23 individual routes with the predictions of the SAFE model for the same routes (**Fig. 4**). The overall correlation was strong (r = 0.88, $P < 0.001$).

## DISCUSSION

The automated method of collecting subjective fatigue ratings was relatively simple to implement and yielded large quantities of data in a nonintrusive way. Furthermore, we found that the automated fatigue ratings at the top of descent responded as expected to factors incorporating crew size, time of day, length of duty, and circadian changes. Of note was that no average scores on any route were above 5.0, which is often taken as a critical value (2), in keeping with the results of previous studies on the same routes. Scores on the daylight "out-and-back" flights were also all lower than those on the long-haul (mostly nighttime) duties. On these out-and-back duties, the mean scores were all lower prior to the first approach than prior to the second approach at the end of duty, so that fatigue was increasing with duty length. The comparison between the different out-and-back duties showed the expected differences based on duty start time, but there was no reduction in fatigue associated with the presence of a third pilot on these duties. This apparent lack of benefit from the additional pilot may be explained by the fact that these duties occurred at times of day when it was unlikely that the pilots would sleep during their rest breaks. This would be expected to reduce the benefit of the extra pilot (3).

The pattern of results on the long haul duties also supported the validity of the collection procedure. Among these duties, the sole daylight sector (Auckland-Tokyo) scored significantly lower than the other duties, which were all operated through the hours of darkness, as would be expected (9). For example, the Auckland-Hong Kong sector was of the same duration and crew complement as Auckland-Tokyo, but departed in the late evening rather than in the morning; the results from the top-of-descent automated ratings showed a significantly higher mean level of fatigue for the Hong Kong night sector than Auckland-Tokyo. We also examined the effect of an additional pilot by comparing the late evening Auckland-Hong Kong three-pilot sector with the late evening four-pilot sector Auckland-Beijing and the early evening four-pilot sector from Auckland-Vancouver. A fourth pilot is only added to mitigate against longer sectors by allowing additional in-flight rest. It is, therefore, not unexpected that there was no difference between the similarly timed three-pilot (Hong Kong) and four-pilot (Beijing) flights; it is likely that the observed difference between the Hong Kong and the

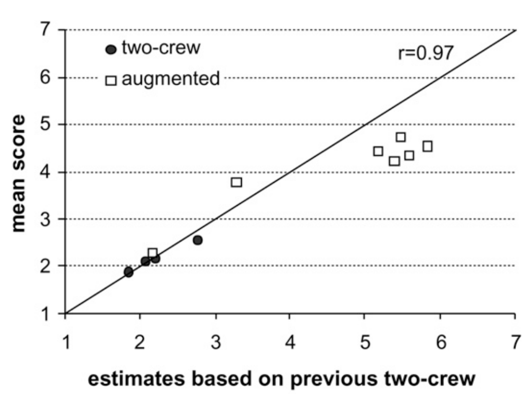


**Fig. 3.** Mean Samn-Perelli scores vs. predictions extrapolated from previous two-pilot results.

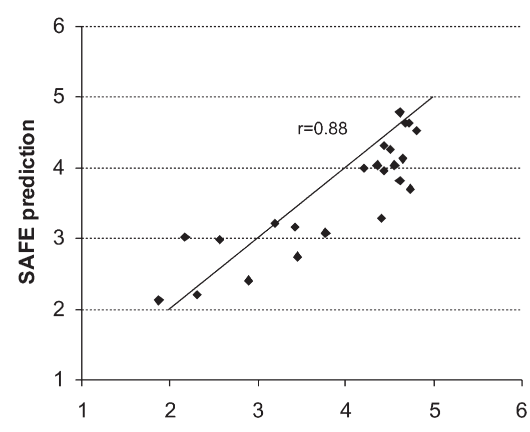


**Fig. 4.** Mean scores vs. predictions of the bio-mathematical SAFE model.

earlier Vancouver flights was related to the different departure times.

We evaluated the validity of the automated collection of pilot fatigue ratings at the top of descent by comparing the results obtained by this method with data from previous studies and to the results predicted by a validated fatigue model. We first compared the results from this study with trend curves derived from a previous "top of descent" study using standard questionnaires. When comparing the data from the current study, it was seen that the effect of an augmented (three or more pilots) crew was to decrease the fatigue level from that expected from the previously published trend curves which were based on two-pilot crews. This is as expected, since the augmented crew arrangement provides opportunities for in-flight rest which are not possible in a standard two-pilot crew (3).

There is increasing use of bio-mathematical models in predicting fatigue in flight operations (4). Although some of these models have been well validated in studies of aircrew, there is a need for continual updating and validation of the models. Our analysis showed close agreement between the outputs of one such model, SAFE, and the top-of-descent data, suggesting that the model and top-of-descent alertness ratings may complement each other. A strength of this study is that it addresses a practical problem faced by commercial airlines: it introduces a methodology which provides a method of gathering subjective data at a critical phase of flight without the need for specialized testing. This makes it possible to evaluate fatigue in a large sample of pilots engaged in an actual airline roster, thereby enabling a more reliable and representative measure of day-to-day operations.

There are some possible limitations to this study. We did not collect information on the work patterns of the pilots prior to the schedules under study. In addition, we have not studied their sleep patterns preflight or in flight. These limitations were an inevitable consequence of the anonymous and brief nature of this method for collecting data. The choice of the time just prior to commencing descent does have a potential drawback relating to in-flight rest: one (or in a four-pilot crew, possibly two) of the pilots may have just returned from bunk rest and it is possible that, despite being asked not to, some pilots undertook the rating when still suffering from the effects of sleep inertia. This could lead to an overestimate of the levels of underlying fatigue at the top of descent in some pilots on long-haul flights. Finally, there was no performance testing to accompany the subjective ratings and the potential, therefore, exists for distortion of the results by some pilots.

However, the large numbers of ratings obtained and the relatively small variability across the data set tend to suggest that the potential for distortion by a few individuals was minimal. A further important benefit of this approach is that it encourages a discussion by the crew of their fatigue and alertness levels just prior to commencing the approach. This enables them to integrate fatigue into the threat assessment when briefing that approach and, thus, offers a direct safety benefit for the operation.

There is potential for further work in this area: in particular, data could be collected at other phases of flight, such as pre-departure. The analysis would be enhanced by amending the software to input automatically the number of pilots present on each flight and potentially by collecting information on the in-flight and preflight sleep history of each pilot. We also believe that these findings have implications for flight-deck design, in which there is a search for better methods of managing fatigue and alertness. Many airlines have introduced flight data analysis programs such as Flight Operations Quality Assurance, which take information from the aircraft data frame on a range of flight path parameters and control inputs; these data are de-identified and analyzed in detail as part of the airline safety management systems. If in-flight alertness data could be integrated into such programs, significant headway could be made into determining the safety consequences of different levels of crew fatigue.

We have demonstrated that onboard ratings at top of descent are a useful method for identifying problem flights and for examining trends across the operation. The data are collected easily, in large numbers, in a non-intrusive fashion. As predicted, the fatigue scores responded as expected to duty length, time zone shifts, and night flying, and correlated well both with previous questionnaire data and with predictions from an aircrew fatigue model. There is potential for the further development and application of this methodology.

ACKNOWLEDGMENTS

*Authors and affiliations:* David M. C. Powell, M.B.Ch.B., FAFOEM, Air New Zealand and University of Otago, Auckland, New Zealand; Mick B. Spencer, B.A., M.Sc., MB Spencer and Associates, Sandhurst, Berks, UK; and Keith J. Petrie, M.A., Ph.D., University of Auckland, Auckland, New Zealand.

REFERENCES

1. Belyavin AJ, Spencer MB. Modelling performance and alertness: the QinetiQ approach. Aviat Space Environ Med 2004; 75(3, Suppl.)A93–103.
2. Civil Aviation Authority. Aircrew fatigue: a review of research undertaken on behalf of the UK Civil Aviation Authority. Norwich, UK: The Stationery Office; 2005. CAA Paper 2005/04.
3. Eriksen CA, Akerstedt T, Nilsson JP. Fatigue in trans-Atlantic airline operations: diaries and actigraphy for two- vs. three-pilot crews. Aviat Space Environ Med 2006; 77:605–12.
4. Gander P, Hartley L, Powell D, Cabon P, Hitchcock E, et al. Fatigue risk management: organizational factors at the regulatory and industry/company level. Accid Anal Prev 2011; 43:573–90.
5. Goode JH. Are pilots at risk of accidents due to fatigue? J Safety Res 2003; 34:309–13.
6. Nicholson AN, Pascoe PA, Spencer MB, Stone BM, Roehrs T, Roth T. Sleep after transmeridian flights. Lancet 1986; 328:1205–8.
7. Powell D, Spencer M, Holland D, Petrie K. Fatigue in two-pilot operations: implications for flight and duty time limitations. Aviat Space Environ Med 2008; 79:1047–50.
8. Powell DM, Spencer MB, Petrie KJ. Fatigue in airline pilots after an additional day's layover period. Aviat Space Environ Med 2010; 81:1013–7.
9. Samel A, Wegmann H-M, Vejvoda M, Drescher J, Gundel A, et al. Two-crew operations: stress and fatigue during long-haul night flights. Aviat Space Environ Med 1997; 68:679–87.
10. Samn SW, Perelli LP. Estimating aircrew fatigue: A technique with implications to airlift operations. Brooks AFB, TX: USAF School of Aerospace Medicine; 1982: Technical Report No. SAM-TR-82-21

**Author Query sheet–ASEM3115**

There are no queries in this article.