Fatigue Safety Performance Indicators (SPIs): A Key Component of Proactive Fatigue Hazard Identification

I. INTRODUCTION

FRMS processes provide many tools that can be used to manage and mitigate organizational fatigue. Metrics are essential to identify potential areas for additional attention and for monitoring the effectiveness of various fatigue management approaches, including Fatigue Risk Management System (FRMS) processes. These metrics, also known as safety performance indicators (SPIs), are reviewed by the Fatigue Safety Action Group (FSAG) to identify fatigue hazards in the operation for risk assessment, mitigation and monitoring. They may also play a role in the monitoring of mitigations and unintended consequences. It is important to look at a variety of metrics rather than relying on a single measure to gain a comprehensive overview of the impact of fatigue within your operation.

The following SPIs are reviewed in this document to assist carriers in developing processes and procedures to monitor the effectiveness of fatigue management approaches:

- Roster Metrics
- Fatigue Reporting Metrics
- Subjective Fatigue Survey Metrics
- Subjective Alertness/Sleepiness Assessment Metrics
- Subjective Sleep/Wake Diary Metrics
- Objective Performance Metrics
- Objective Sleep Metrics
- Fatigue Model Metrics

The above described metrics may be collected in many ways. These include subjective surveys, such as daily diaries including sleep logs and work schedules as well as objective data collection tools such as actigraphy and objective performance tests. The most meaningful information is often produced by collating and comparing data collected through a combination of techniques.

The specific metrics selected should be identified at the time that the control or mitigation, as part of the FRMS, is implemented. These metrics should be tracked over time to assess their effectiveness. The best metrics for assessing mitigations vary and should be selected with an understanding of the operation and intent of each mitigation. Extracts from the *ICAO Implementation Guide for Operators* regarding metrics is also included at the end of Section II.

For all of the metrics mentioned described in this document, the operator may wish to further review the data according to:

- Fleet
- Base
- Rank
- Operations type (e.g., passenger versus cargo)
- Sector/city pair

This document also presents an overview of the types of data collection studies that can be employed to collect fatigue SPIs and summarizes some of the major issues that should be considered during analyses. The information can be used by carriers for guidance in planning and conducting Proactive Fatigue Hazard Identification with their operations.

II. SAFETY PERFORMANCE INDICATORS (SPIs)/METRICS

a. Roster Metrics

Roster metrics are generated based on data in the rostering software. The rostering software is designed to facilitate roster building, roster delivery, and the storage of rosters for legal purposes.

Potential roster metrics include:

- Captain's Discretion - percentage of duties
- Total and average flight hours per month
- Total and average duty hours per month
- Number of minimum rest periods per month as a percentage of all rest periods
- Number of unfit for duty due to fatigue resulting from previous roster as a percentage of all duties
- Number of extended flight duty periods for a specific pairing
- Percentage of sectors on which controlled rest was used
- Reserve callout as a percentage of all duties and monthly utilization
- Overtime usage

Detailed analysis of roster data may also provide insights into roster stability and other factors that may contribute to elevated levels of fatigue risk.

b. Fatigue Reporting Metrics

Crew member reports about fatigue-related challenges, events and incidents are critical to keeping the carrier informed about fatigue hazards in day-to-day operations within their specific operations. Effective fatigue reporting systems must be implemented within an effective reporting culture. Protective measures that protect the confidentiality and anonymity of the data need to be in place. Individuals who have access to the reporting system should clearly understand how the data is collected, why the data is being collected, how it will be analyzed, and how the results and corrective actions will be implemented within the organization. Compliance with submitting fatigue reports will potentially be higher and more beneficial to the operator with the use of a fatigue report form (FRF) that is clearly worded, easy to access, easy to complete, and easy to submit.

Fatigue report forms (FRF) should include information about previous sleep and duty information for a minimum of 72 hours prior to the event or incident, the time of the event, commuting details, subjective ratings, and any other fatigue related aspects related to the event or incident. Providing an open field for the individual to report other details in free text form is also helpful. The fatigue reporting metrics available depend on the data that is collected via the fatigue report form.

- Examples of fatigue reporting metrics include: Number of FRFs completed and submitted, expressed as a percentage of crew sectors
- Breakdown of FRFs by reason for submission (e.g., adverse event, unfit for duty due to fatigue, controlled rest, general fatigue concern)
- Breakdown of FRFs by sector
- Subjective fatigue scores (e.g., Samn-Perelli) at sign-on/time of day/at the time of the adverse event
- Average sleep achieved in the prior 24/48h/72h and average hours of wakefulness
- Common symptoms of fatigue reported

c. Subjective Fatigue Survey Metrics

Carriers can learn a great deal about perceived fatigue within their organizations, and about the factors that may contribute to fatigue, by conducting subjective fatigue surveys within their workforce. Administering such surveys is a very effective method for quickly collecting a large quantity of information from a large population. For example, a survey might ask people to evaluate potential signs of fatigue of coworkers, identify the duties that they personally find the most fatiguing, identify specific city pairs or schedules associated with increased fatigue, or identify other operational variables that may be contributing to fatigue. The data can be analyzed through a variety of sophisticated statistical techniques to identify patterns of fatigue among survey respondents and to identify other challenges within operations that might be appropriate for further objective evaluation or immediate mitigation.

Survey questions should be developed to target specific areas of concern or to gather general information concerning fatigue within the specific aviation organization or population. Information should be collected across a wide-range of fatigue-related topics. Typical fatigue-related surveys include:

- Demographic questions (What is your age, gender, etc.?)
- Questions about perceived fatigue (e.g., How fatigued do you feel after a day at work? Do you see co-workers that are fatigued?)

- Fatigue awareness questions (e.g., What techniques are most effective at managing fatigue?)
- Population – specific questions (e.g., Where do you live, work, etc.? How long is your daily commute?)
- Questions about habit and lifestyle (e.g., How many hours do you typically sleep each night?)
- Questions about perceived problems (e.g., At what time of day are you typically most challenged by fatigue?)

Best practices that are useful to observe when planning and administering surveys include the following:

- Include survey participants (i.e., target population such as flight crew, cabin crew, maintenance, flight dispatchers, etc.) in the formulation of questions to help ensure that questions are appropriate for the specific group being studied and easily understood by the group who are completing the surveys.
- Assure participants that all collected data will be confidential, cannot be traced to respondents and will be non-punitive.
- Communicate about the survey clearly and frequently; stress the long-term benefits of the survey and include links to the survey in all correspondence.
- If the survey is to be conducted online in an environment where some individuals may not have sufficient access, it is important to provide alternative strategies for participation.

d. Subjective Alertness/Sleepiness Assessment Metrics

Subjective alertness/sleepiness assessments are valuable tools for proactively assessing the level of fatigue within aviation operations, or among a particular working population. They can help to identify specific aspects of work schedules or other factors that may contribute to fatigue. Repeated assessments may also provide a basis for assessing whether and how effective workplace changes, policies, or other interventions implemented have been in managing and mitigating fatigue.

Subjective alertness/sleepiness assessments typically ask individuals to rate their fatigue, sleepiness, or alertness levels at a given moment in real time. For example, individuals are asked to rate their alertness/sleepiness at sign-on/check-in/report for duty, TOC, before and after each sleep period, TOD and sign-off/check-out/end of duty.

These surveys tend to be short for very quick assessments. They can be administered at multiple times during the day or across a duty cycle to assess how fatigue, alertness or sleepiness changes across the day or duty period. It is important that these surveys have been scientifically validated to ensure that the data being collected is reliable and measures the variables of interest at all times of day and at varying levels of sleep loss.

Validated tools for assessing subjective alertness/sleepiness include the Karolinska Sleepiness Scale (KSS) (Akerstedt & Gillberg, 1990), the Samn-Perelli scale (Samn & Perelli, 1982), the Stanford Sleepiness Scale (SSS) (Hoddes et. al, 1973), and the Profile of Mood States (POMS) (McNair et. al, 1971). For the POMS, the items "confused," "fatigued," and "vigorous" are most closely related to measures of objective sleepiness (and, in the case of "vigorous," objective alertness). Visual analog scales (VAS) also have been used for subjective alertness/sleepiness assessment.

e. Subjective Sleep/Wake Diary Metrics

Daily sleep/wake diary data can provide insights into the actual work habits and sleep/wake schedules of individuals. They can also provide a basis for modifying schedules, providing training, or implementing other interventions. Individuals are asked to record their time in bed, time awake and other sleep-related factors. Often, they are asked to subjectively rate their sleepiness and alertness before and after sleep periods, rate their sleep quality, indicate the requirement for more sleep, and the sources of any sleep disturbances.

f. Objective Performance Metrics

Objective performance assessments within the operational setting can be conducted by means of a standalone testing device. These measurements provide objective data that can be used to supplement the subjective data collected in subjective surveys and assessments. It allows for measuring fatigue-related changes in performance. The most widely accepted approach for collecting objective performance measures during operations is the use of simple tests that were developed and validated in the laboratory to be sensitive to sleep loss and fatigue. Stand-alone assessments conducted in operational environments are typically administered by means of a handheld device such as a personal digital assistant (PDA) or smartphone (Roma et.al, 2010; Rosekind et. al, 2006). These short performance tests can be administered at various times that are associated with pre-flight duties, TOC, TOD, post-flight, pre-bunk sleep, post-bunk sleep, etc.

One of the most commonly used and validated test that has been used in both operational and laboratory-based research studies is the Psychomotor Vigilance Test (PVT). The PVT possesses desirable properties including (a) no learning curve (i.e., no "practice" effect) and (b) demonstrated sensitivity to sleep loss in a variety of different environments and populations (Dinges & Kribbs, 1991; Dinges & Powell, 1985). Measuring an individual's 3 or 5 minute PVT test at various times during the day provides a measure of their performance at those times. The specific timing for assessments can be varied according to the focus of the study and the needs of the carrier.

g. Objective Sleep Metrics

Objective sleep measurements can be recorded using actigraphy or measuring activity levels using a wrist watch like device. Actigraphy is a well-accepted, unobtrusive and readily-available technology for the quantification of sleep (Stone & Ancoli-Israel, 2011). Activity levels recorded by a wristwatch-like actigraph are interpreted by an algorithm that provides an estimate of sleep duration. These estimates are considered reliable as they correlate very highly with the estimation of sleep as performed by polysomnography - the gold standard for sleep measurement. Analyses of the actigraphy data allows for the calculation of sleep duration, bedtime, wake time and sleep quality.

h. Fatigue Model Metrics

The specific metrics available will depend on the fatigue model in use, but all models involve setting a threshold and focusing on operations predicted to be outside the threshold. The strength in modeling lies with their ability to model multiple physiological systems that regulate fatigue and provide an estimate of fatigue risk, which can be expressed in terms of alertness, performance, risk level, sleepiness, etc. The results can then be used for relative comparisons of potential fatigue risk associated with the sector. This information can then be used to develop schedules and optimize pairings that promote alertness and mitigate fatigue-related risk. There are many factors that affect one's fatigue level; modeling metrics are just one part of the equation and should always be used in combination with other data and information that is available.

Potential metrics include:

- Percentage of all duties outside the threshold
- Percentage of critical phases of flight outside the threshold e.g., take off, approach, and landing
- Ranking of sectors by predicted fatigue outside the threshold
- Identifying common factors for sectors with predicted fatigue outside the threshold

4.6 FRM Processes Step 5: Risk Mitigation[1]

The effectiveness of implemented controls and mitigations must be assessed, which requires setting safety performance indicators such as the following.

Schedule-related indicators:

- Number of flight deviations (or flight completion not accomplished) on specific city pairings, due to fatigue, lack of staff, medical emergencies, etc.
- Number of bids for pairings identified as high fatigue risk (e.g., back-to-back night flights).
- Number of flight duty period exceedences into allowable excesses (as determined through risk assessment. For example, longer than 14 hours.)
- Number of flight duty periods determined to be "significantly" greater than scheduled.
- Number of flight duty periods longer than a specified number of hours without a rest break within the duty.
- Number of flight duty periods starting or finishing within the window of circadian low (WOCL).
- Number of take offs and landings within the WOCL.
- Number of duty periods with more than a specified number of flight sectors.
- Number of duty periods with more than a specified number of aircraft changes.
- Number of successive early wakeups, especially combined with long "sits" between flights or long duty days.
- Number of reduced rest breaks within duties (by more than a specified number of minutes determined to be "significant").
- Number of reduced rest breaks between duties (by more than a specified number of minutes determined to be "significant").
- Number of reserve crew call-outs (on particular flights, at a particular crew base, etc.).

[1] Fatigue Risk Management Systems for Regulators, ICAO Doc 9966, 2012 Edition, page 4-18

Proactive/reactive fatigue indicators:

- Measured data outside acceptable thresholds (e.g., sleepiness ratings, Psychomotor Vigilance Task (PVT) scores, or inadequate layover sleep duration).
- Number of fatigue reports (sorted in many ways such as by crew base, seat, augmented flights, fleet types, operational types, etc.).
- Number of fatigue-related incidents.
- Number of fatigue-related FOQA events associated with a particular schedule for which fatigue reports have been received.
- Absenteeism/fatigue calls.

5.2 FRMS Safety Assurance Processes Step 1: Collect and Review Data[2]

Examples of safety measures/metrics obtained through the FRM processes:

- The number of exceeded maximum duty days in operations covered by the FRMS;
- The number of voluntary fatigue reports per month;
- The average "fatigue call" rate by flight crews on a specific pairing (trip);
- The ratio of fatigue reports from ULR operations covered by the FRMS to fatigue reports from the long haul operations covered by the prescriptive flight and duty time regulations.
- The level of crew member participation in fatigue-related data collection;
- The number of times fatigue is identified as an organizational factor contributing to an event.

[2] Fatigue Risk Management Systems for Regulators, ICAO Doc 9966, 2012 Edition, page 5-4

III. DATA COLLECTION PROTOCOLS

Data collection protocols that use a combination of both subjective and objective metrics provide meaningful data that can be used as the foundation for informative Proactive Hazard Identification studies. While subjective studies can provide valuable information, they are limited to collecting data based on self-assessment and observation. Overreliance on subjective study data on its own can be misleading. Similarly self-reports of seemingly objective information, such as hours of sleep or time spent at work, are often based on inaccurate recollections and should be treated with caution.

Measurement tools such as actigraphs and reaction time tests can provide objective measures of activity levels (from which sleep amounts can be derived) and performance. However, sometimes the data collected by these tools in operational environments can be difficult to interpret in isolation. Therefore, studies that combine subjective data capture techniques, such as surveys or daily logs with objective assessments, can provide meaningful data.

a. Subjective Survey/Questionnaire Study

Survey/questionnaire studies provide a database of information that can be analyzed in many ways. These retrospective surveys are typically very long and ask individuals very detailed information about their prior work history, sleep history and experiences with fatigue and sleepiness. The questions ask the respondent to provide information for a period of time that has occurred in the past and can be conducted before and after a control or mitigation is implemented. The use of standardized fatigue surveys allow the carrier to compare fatigue variables between and/or across operations and with other data collection studies.

Results

Specific analyses techniques depend on the questions being asked or the goal of the survey. Examples of statistical findings that might result from survey analysis include:

- X% of flight crew rated xxx-xxx as the most fatiguing sector or pairing as applicable
- X% of the people who work on Y-type of flights report getting 6 or less hours of sleep prior to commencement of duty
- The most commonly reported fatigue aspects in the operation were x, y, & z

b. PDA Study

The "gold standard" for Fatigue Hazard Identification studies is one that combines subjective data capture with objective measures of sleep and performance. Use of daily diary information, subjective estimates of alertness and sleepiness, objective performance, and other fatigue related measures in combination with actigraphy is a validated methodology for assessing potential fatigue risk in aviation operations.

This is known as a PDA study (Psychomotor Vigilance Test, diary, actigraph) and includes the following components:

- Objective sleep metrics
- Objective performance metrics
- Subjective sleep/wake diary metrics
- Subjective alertness/sleepiness metrics

Results

By analyzing data from multiple sources, PDA studies can provide data for analysis that often leads to insights regarding fatigue mitigation and management within the organization. Some of the statistics that can be obtained include:

- Sleep and wakefulness: sleep duration, duration of wakefulness, sleep start time, sleep finish time, sleep quality, requirement for more sleep
- In-flight sleep: in-flight rest schedule (duration and timing of breaks), amount of sleep obtained during breaks, influence of environment on amount of sleep obtained (bunk/crew compartment versus seat, noise, turbulence, temperature, light, etc.)
- Self-rated sleepiness: pre- and post-sleep sleepiness, in-flight sleepiness (including comparisons of subjective reports and objective assessments)
- Crew performance: mean response time for all trials on PVT

c. Top of Descent (TOD) Study

In these studies subjects rate their fatigue at TOD (or predicatively/retrospectively if TOD is not a suitable time for data collection). Additional data collected on a TOD study could include rating fatigue at start of duty to enable fatigue acceleration/deceleration calculations (difference between start of duty and TOD fatigue ratings). Rating fatigue from prior duties enables the identification of sectors that are not fatiguing in themselves, but may be fatiguing due to previous sectors worked.

Results

Statistics that can be calculated as a result of TOD studies include:

- Fatigue at TOD by sector – rate all sectors
- Flight duty periods with a high fatigue rating at the start of duty
- Sectors with abnormal accumulation of fatigue; e.g. large difference between start of duty fatigue ratings and fatigue ratings at TOD or end of duty

d. Daily Diary Study

Diary studies are generally designed to collect information about work schedules, sleep schedules, and alertness and fatigue ratings. In some case, other fatigue-related variables may be collected. Daily diary studies can provide insights into the actual work habits and sleep schedules of individuals, and can provide a basis for modifying schedules, providing training, or implementing other interventions. Information on in-flight rest and sleep periods can also be collected with the daily diary. Individuals record the duration and timing of in-flight rest, duration and timing of in-flight sleep, describe and assess the sleep environment (bunk/crew rest compartment, first class seat, business class seat, etc.), and list the sources of any sleep disturbances that they may have experienced (noise, turbulence, temperature, light, etc.).

Examples of the types of information that is collected during a daily diary study include:

- Subjective sleep measurements
- In-flight sleep
- Subjective alertness/sleepiness measurements

Results

Statistics that can be calculated based on daily diary data include:

- Sleep and wakefulness: sleep duration, duration of wakefulness, sleep start time, sleep end time, sleep quality, requirement for more sleep

- In-flight sleep: in-flight rest schedule (duration and timing of breaks), amount of sleep obtained during breaks, influence of environment on the amount and quality of sleep obtained (bunk/crew rest compartment versus seat, noise, turbulence, temperature, light etc.)
- Self-rated sleepiness: pre- and post-sleep sleepiness, in-flight sleepiness
- Self-rated alertness: pre- and post-sleep alertness, in-flight alertness; at sign-on/check-in/report for duty, TOC, before and after each sleep period, TOD and sign-off/check-out/end of duty

Other Indicators of FRMS Effectiveness

Other data may also be used to assess the effectiveness of the FRMS. For example, Corporate Fatigue Culture Surveys can provide insights into how the FRMS is perceived and how the organization behaves with regard to fatigue in general. Comparing the findings from surveys or studies conducted on a recurring basis can provide the basis for analyzing changes to attitudes and behaviors and other trends.

References

Akerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. *Int J Neurosci* 1990; 52(1–2): 29–37.

Dinges DF, Kribbs NB. Performing while sleepy: effects of experimentally induced sleepiness. In: Monk TH, editor. *Sleep, Sleepiness, and Performance*. Oxford, England: John Wiley and Sons; 1991: 97–128.

Dinges DF, Powell JW. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavioral Research Methods: Instruments and Computers* 1985; 17: 652–5.

FRMS Implementation Guide for Operators, 1st Edition, July 2011, IATA, ICAO, IFALPA

FRMS Manual for Regulators, ICAO Doc 9966, 2011 Edition

Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. *Psychophysiology* 1973; 10(4): 431–6.

McNair DM, Lorr M, Druppleman LF. *EITS Manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Test Services; 1971.

Samn SW, Perelli LP Estimating aircrew fatigue: A technique with implications to airlift operations. Brooks AFB, TX: USAF School of Aerospace Medicine. Technical Report No. SAM-TR-82-21; 1982.

Roma PG, Mallis MM, Hursh S, Mead AM, Nesthus T. *Flight Attendant Fatigue Recommendation II: Flight Attendant Work/Rest Patterns, and Performance Assessment*. DOT/FAA/AM-10/22. Washington, DC: Office of Aerospace Medicine, Department of Transportation; 2010.

Rosekind MR, Gregory KB, Mallis MM. Alertness management in aviation operations: enhancing performance and sleep. *Aviat Space Environ Med* 2006; 77(12): 1256–65.