A Best Practice for Quantification of Fatigue Risk

Understanding and improving upon the situation for crew fatigue risk in an operation is of course greatly simplified with a well thought-through way of quantifying this risk. To some extent one can use fatigue reports, collected data and crew feedback after-the-fact, but what is the best practice metric for the fatigue risk in upcoming crew pairings and rosters to operate next month?

Defining Fatigue Risk
There is no formal definition of fatigue risk set by ICAO or IATA. A proven useful definition when planning crew members is: the risk of crew performing a lapse, slip, mistake or violation, negatively impacting flight safety, as an effect of low levels of alertness.

With this definition, the focus thus lies on flight safety and human error primarily among pilots on active flights, rather than ’crew comfort‘ or sleepiness during commute, ground duties or a deadhead flight.

A metric for one flight
Looking at a single flight it is clear that the potential for human error negatively impacting flight safety is greatly elevated during approach and landing, which is the phase of flying most taxing on pilot capabilities. During this time, the workload is normally at its highest and there is little margin for slowing down or double checking oneself or a colleague in order to reduce risk. The consequence of a slip, lapse, mistake or violation is also potentially disastrous. A vast majority of fatigue-related accidents in aviation are related to human error during this phase of flying.

For these reasons it makes sense to focus a metric on estimating fatigue risk primarily using the predicted level of alertness (or sleepiness) close to the end of active flights. An often used point in time for collecting data is close to top of descent (TOD) - making it a good choice for predicted alertness level to represent a flight.

Risk vs. Sleepiness
The risk of a lapse, slip, mistake or violation for an individual has been shown to accelerate as sleepiness increases. Figure 1 illustrates the development in the probability of an accident (dotted line) in a driving simulator where an inflection point is seen just above where subjects are experiencing KSS 8.

However, when predicting future sleepiness, a fatigue model will have limited accuracy for one individual, due to a number of reasons; the models are not perfect, models are under-informed, and there are significant inter and intra individual differences among crew. Figure 2 illustrates how the odds-ratio for an actual accident develops as a function of predicted sleepiness from a bio-mathematical model.

The conclusion to draw, is that a predictive metric capturing fatigue risk should also include a risk contribution from much lower levels of predicted sleepiness than those close to, or passing KSS 8. Human physiology, when being predicted into the future, does not have any sharp ‘thresholds’ separating safe from unsafe. The probability of an accident accelerates more slowly, and from lower levels, when sleepiness is predicted, compared to the risk development observed for self-assessed sleepiness. Figure 3, further below, based on FDM data, tells a similar story.

A metric for a set of flights
The focus of fatigue risk management when scheduling crew should of course be to reduce the overall risk for the operator to suffer an incident or accident.
What is really achieved if we reduced fatigue risk on, say, the twenty worst flight duties, if the system response from those changes is negative with the overall risk increasing? [Example here below.]

**Example**: A reduction of ‘maximum duty time’ for overnight flight duties may seem as a great idea at first for reducing fatigue risk. However, the flights are still present in the flight schedule and will need to be flown. The modified rule may lead to the creation of a lot more night duties, each with commute time before and after, potentially inflicting on physiologically sound timings for sleep. The change will create a need for scheduling more consecutive night duties, stacking up sleep debt for crew. What was perceived as an improvement when looking at ONE night duty in isolation, may well have been the exact opposite looking at the overall operation. This system response from changes made is far to often overlooked.

For this reason, it is crucial to have methods for quantifying, tracking and controlling overall fatigue risk, using a metric that adds up all small probabilities for the individual flight assignments, rather than working with flights in isolation. When doing so, it is logical to use a weighted sum over the set of flights with a ‘weight’ that accelerates when the predicted sleepiness increases, reflecting how fatigue risk develops in individuals (figure 2 and 3 again).

There is no formal standard for this, let alone a way of practically establishing the optimal shape for such a weighting function. Even so, not allowing ‘perfect to become the enemy of good’, we can approximate a shape that mimics the acceleration of risk we do know exists. At Jeppesen, a simple quadratic shape for the risk contribution is used, accelerating from KSS 5 and assigning risk contribution to all flights above that level. Our scale is, however, the other way round as BAM is predicting alertness on the Common Alertness Scale (CAS) from 0 to 10,000, which is anchored to the KSS scale, going in the opposite direction.

The overall risk metric has been named AFR, for **Absolute Fatigue Risk**, and serves as a proxy for the overall probability of an incident of accident. The higher the number, the higher the risk.

\[ AFR(x) = \begin{cases} (50 - x)^2 & x < 50 \\ 0 & \text{otherwise} \end{cases} \]

where \(x\) is the predicted alertness in CAS divided by 100

Using AFR

AFR is now our best practice metric reflecting overall fatigue risk in a set of flights. It takes both frequency (number of flights) and the severity into account and can be used in a number of helpful ways when planning crew:

- Quantify the ‘system response’. In our example, will the shortening of maximum duty time result in overall lower risk? Do we also need to limit consecutive night duties?
- Suppress risk during crew pairing and roster optimization by allowing AFR to feed into the objective function.
- Distribute fatigue risk among crew, sharing the burden
- Track the risk development over time
- Direct focus to the right part of the operation
- Use as a risk ‘profiler’ by dividing AFR over the number of flights; a metric named NFR (Normalised Fatigue Risk)

Figure 4 shows how the same set of flights have been planned in two different scenarios but with a clear difference in risk. Both scenarios respect the same planning rules but we can, by just looking at the distributions, quickly confirm that scenario B is to prefer as it contain much lower risk. (Fewer flights in the left tail of the distribution). Our AFR and NFR metrics are confirming the same thing but also quantifying that the risk has been reduced by 45%.

The AFR/NFR approach described is today used by a large number of Jeppesen customers to control and reduce overall fatigue risk and is an already established best practice, allowing them
to reach much further in risk management compared to only limit predicted sleepiness/alertness/fatigue/effectiveness with a sharp cut-off at a certain level (of little or no real importance).

The NFR metric turns out to be very useful also when comparing the risk profile between different fleets, bases, ranks, destinations, and even airlines and for tracking risk development over time. Table 1 lists some typical NFR ranges for various types of operations.

Whereas AFR varies with the volume of the operation, NFR is ideal for benchmarking and has now overtaken the position ‘PA5’ had a few years ago [3] as the main risk profile metric. NFR may be somewhat less intuitive, but has the advantage of working well also on small sets of flights, like a pairing or one roster, plus reflecting fatigue risk development in human physiology in a more detailed way.

<table>
<thead>
<tr>
<th>NFR</th>
<th>Type of operation / business model / risk appetite</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;400</td>
<td>Long-haul cargo operations with multiple crew bases in different time zones, with sub-standard FRM practices or rules dealing with TZ transitions.</td>
</tr>
<tr>
<td>250-400</td>
<td>Challenging long-haul operations such as charter operations with two pilots with multiple TZ crossings and short out-station layovers. Also some 24/7 short haul PAX operations in South and Mid America, the gulf region and Russia, as well as some domestic night cargo operations.</td>
</tr>
<tr>
<td>100-250</td>
<td>Challenging short-haul operations, and mid-haul, with a fair amount of back-of-the-clock operation requiring stacking up consecutive late (or night) operation.</td>
</tr>
<tr>
<td>50-100</td>
<td>Normal short-haul PAX operation, performed for example by many European and North American flag carriers.</td>
</tr>
<tr>
<td>&lt;50</td>
<td>Predominantly day-time PAX operation with good FRM practices.</td>
</tr>
</tbody>
</table>

Table 1. Typical NFR (risk profile) values for different operations. [4]

Further reading:

- Aligning Rules With Human Physiology
- Are your processes in control?
- Jeppesen Concert
- Assignment-centric Performance Indicators
- BAM Safety Performance Indicators
- The secret behind pro-active risk reduction
- Your opinion is interesting, but...

[4] Approximate ranges only. An individual operator may deviate from ranges stated.