Measuring Operational Safety in Aviation

Abstract

The drive for operational efficiency in aviation means that aircraft operations are increasingly run against a backdrop of measures and targets. This in turn generates an increasing need and desire to include safety as a metric that can be tracked and monitored. Safety however is not something that can be measured directly but risk has the potential to be an indicator of safety.

The risk rating of individual incidents and reports provides a data set that has the potential to provide an overall indicator of operational risk. That data and its risk assessment is subject to a number of limitations, not least the fact that at best reporting data can only give a partial view of the risks that the operation faces. These limitations mean that any measure based on this data is fragile in nature and has to be presented and treated accordingly.

Presenting the data to show the overall risk trend for the organisation can potential lead to negative pressure on the reporting rate. To avoid this the reporting rate needs to be an integral part of any risk trending to ensure the value of increased reporting is recognised.

Ultimately the greatest value in collecting safety data is to use it to reduce the likelihood of an accident or serious incident and therefore any measures derived from it should have this objective. Risk measures can be used to provide a focus on the areas of the operation with the most risk and enable effort to be applied appropriately in those areas. The same measures will then allow you to monitor the effectiveness of those actions by the subsequent reduction in risk.

Measuring Safety

The ever-increasing drive for operational efficiency ensures that aircraft operations are run against a backdrop of measures and targets designed to ensure that the operation is being managed and run efficiently. With such a data driven operation, to keep safety in the top priority, the desire to measure and track safety is strong. Where accidents occur often, the measure of safety could potentially be a simple task of measuring their frequency, however the absence of accidents does not necessarily imply a high level of safety. Obviously we have to look further as waiting for an accident to determine you are not safe is clearly not an acceptable way to manage an aircraft operation.

We therefore need to provide a measure of safety that does not rely on accidents, but can we effectively and reliably measure operational safety? Safety has been described as a construct and a concept (Vick[1]) and as such is clearly not a quantifiable entity. Safety is much more of a personal judgement than a finite measure and therefore does not lend itself to being reliably measured. The US Federal Aviation Administration describes 'Safety Risk' as a measure of probability and impact, leading us on to the concept of risk. Risk is something that the industry does try to quantify.

Measuring Risk

It is not the intent of this paper to address the whys and hows of the assignment of risk ratings, nevertheless it is important to look at some of the key concepts and difficulties that it involves. Fundamentally, to measure the risk of something you must define the threat to which you wish to rate it against and then you must make an assessment against that threat, neither of which is straightforward. Clearly in aviation the key threat is that of an aircraft accident, but other threats such as personal injury and damage are also areas that managers will want to measure and track.

Once the threat is defined an assessment must be made of the event or activity against that threat and a rating given. How do you make that assessment? Is it a formal, scientific process that provides a repeatable result? If the risk assignment is based purely on the severity of the event and its frequency as many processes do (Macrae et al[2]), then it potentially could be a scientifically repeatable process. Unfortunately the actual severity of an event is not a good measure of its risk. For example, consider the injury risk of a baggage cart rolling over an employee's foot compared to an aircraft narrowly avoiding, by luck, rolling over an employee. The first involved actual injury and therefore greater actual severity than the potential occurrence in the second, but the second event is clearly of greater risk.

Therefore an effective rating of risk must include, either formally or informally, an element of potential severity. However ‘potential’ is often in the mind of the beholder and therefore the rating of risk becomes subjective and will clearly vary over time and between people.

The subjectivity of any risk rating is clearly one issue, another is the purpose for which the assignment of that rating is done. The assignment of risk to an incident could be for the purposes of tracking and trending that risk over time or to determine what immediate action should be taken in response to the incident (the higher the rating the greater the response). The process is likely to include elements of each, further clouding the data that you have.

Limitations of Risk Measures

Not withstanding the caveats and limitations described above, there is potentially a source of quantified risks that could be used to measure operational risk. To produce that overall measure what risk data do you use? The most obvious source is that of reported incident data: Air Safety Reports and other incident report forms. The problem is that risk measures based on such reports can only be as good as the reporting culture of organisation and the employees that raise them.

Even an established and mature reporting culture will still only give you a view of some of the incidents and events that are occurring in the operation. Naturally some employees will report more than others and different attitudes within the same organisation will vary what people report. This difference in reporting will be most significant where a culture of personal blame exists in response to incidents (Rose[3]). Furthermore once overall measures of risk to the operation are based on reporting data, a drive to reduce the ‘perceived’ risk in the operation could be achieved by reducing the quantity of reports. In effect you are only seeing, and now measuring the risk of, what people choose to report.

One way to visualise the problem is to consider a variation of the iceberg model of threats to an organisation. You see the significant events as they are above the waterline and cannot be hidden, however this is just the tip of the iceberg, much of your threat lies below the waterline and is not obviously visible (figure 1). A good reporting culture helps you to raise the iceberg in the water to gain a better view of the total threat, but as suggested above it will never be perfect. In the measurement of the risk you can obviously only rate the part of the iceberg you can see and by using it as an overall measure of risk or safety, you may create a fallacy of false precision (Smithson[4]). You are basing your measure on what you can see (i.e. what is reported) but do not know what you cannot see (i.e. what is not reported). You therefore cannot be sure which is the largest variable and how the variables change over time (figure 2), effectively ‘you don’t know what you don’t know’.

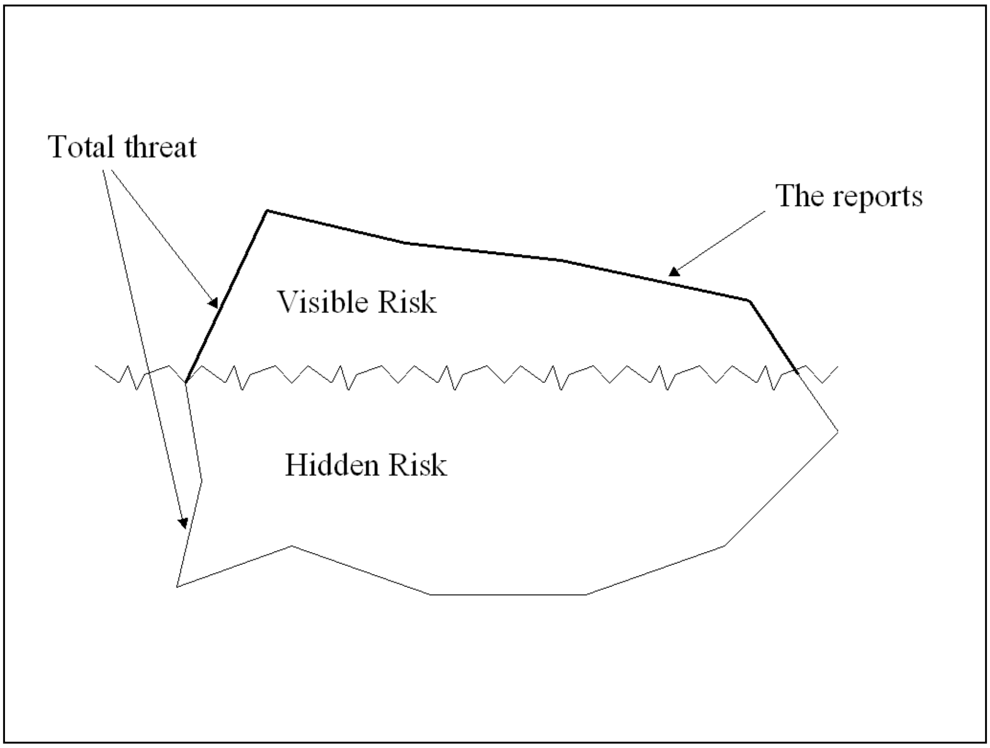

Figure 1 – Reporting Iceberg

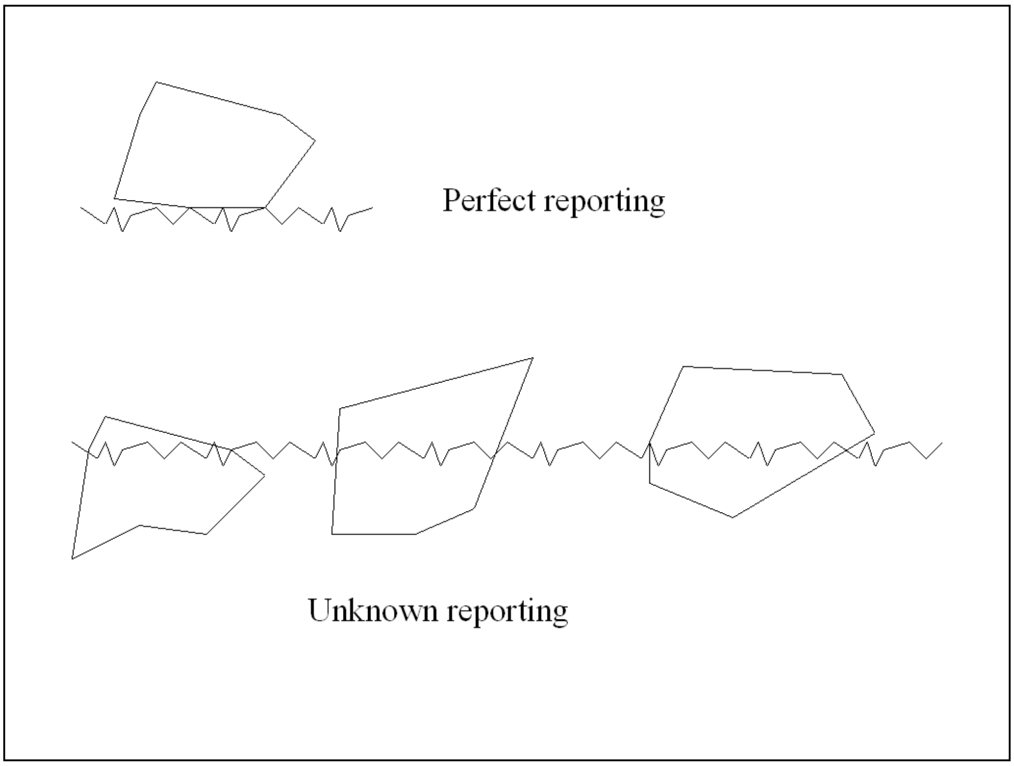

Figure 2 – Variation of the Reporting Iceberg

It is not only reporting culture that drives variations in reporting rate: both procedures and technology do too. A prime example of a procedure driven change would be something like a campaign to stop injuries from overweight bags. The focus on the issue and request for data could generate a spike in reporting of events that were only seen rarely before, implying an increased risk. The introduction of TCAS (Traffic Alerting and Collision Avoidance System) is a good example of a technological driven shift in perceived risk. The introduction of the system produced multitudes of reports of incidents of reduced separation, implying that the risk of midair collision had greatly increased. Ironically both of these spikes are really indicators of decreased risk, as you now better understand the problem and can therefore address it.

There are two potential ways to assist in overcoming these limitations; firstly by using more reliable data and secondly by trying to verify the data to determine how much of the iceberg you can actually see.

Availability of more reliable data about safety in the organisation is limited as it could only come from real-time monitoring of the operation to detect incidents and events as they occur. In aviation one prime source of such data could be from a Flight Data Monitoring Programme, which potentially captures a significant portion of the events that are occurring in the pilot environment. Obviously it is limited to capturing the events that it can detect, but what it can detect is known and all of those events will be seen (depending upon the data replay rate). The key limitation is that there is very little similar data outside of the flight operations environment and many of your risks will lie beyond its reaches. Despite these limitations it is clear that flight data monitoring should not be overlooked when trying to measure operational safety.

The alternative is still to use reporting data but to try to verify that what is reported is a representative and reliable view of the actual events that are occurring in the operation. For example using Flight Data Monitoring to verify the proportion of known events that get reported is one option, another might be to observe activities and assess what should and did get reported. By introducing methods of data set verification the limitations of the data set can be better understood and addressed.

Using Risk Measures to trend performance

Once you have an overall measure of operational risk what can you use it for? Clearly with the limitations and caveats that surround it you have a measure that is nearer sand than concrete and therefore you have to be careful what you base upon it. The data will not be reliable enough to decide that the operation is so safe that there is no need to apply effort to maintain safety; therefore the measure does not lend itself to targets and ‘traffic light' type indicators. Safety has to be considered a process of continuous improvement; the absence of accidents is no guarantee that they will remain absent. Also many of the limitations in the data could be directly driven from the use made of it and therefore the measures must not be used in a way that drives poor behaviour in safety reporting or data management.

As stated previously, if the measure of operational risk is based on reporting data then reduced reporting appears to lower the risk. To avoid the data driving the wrong behaviours in the reporting of incidents, or in the management of safety data, the activity must be focussed away from comparison. Safety is not a competition and the measurement of operational safety must not make it one. There are always two ways to win a competition and in something as complex as maintaining a safe aircraft operation the easiest short-term answer may be to manipulate the data. By avoiding competition and by treating upward reporting trends as a good thing, both the desire and ease of manipulating the data are reduced.

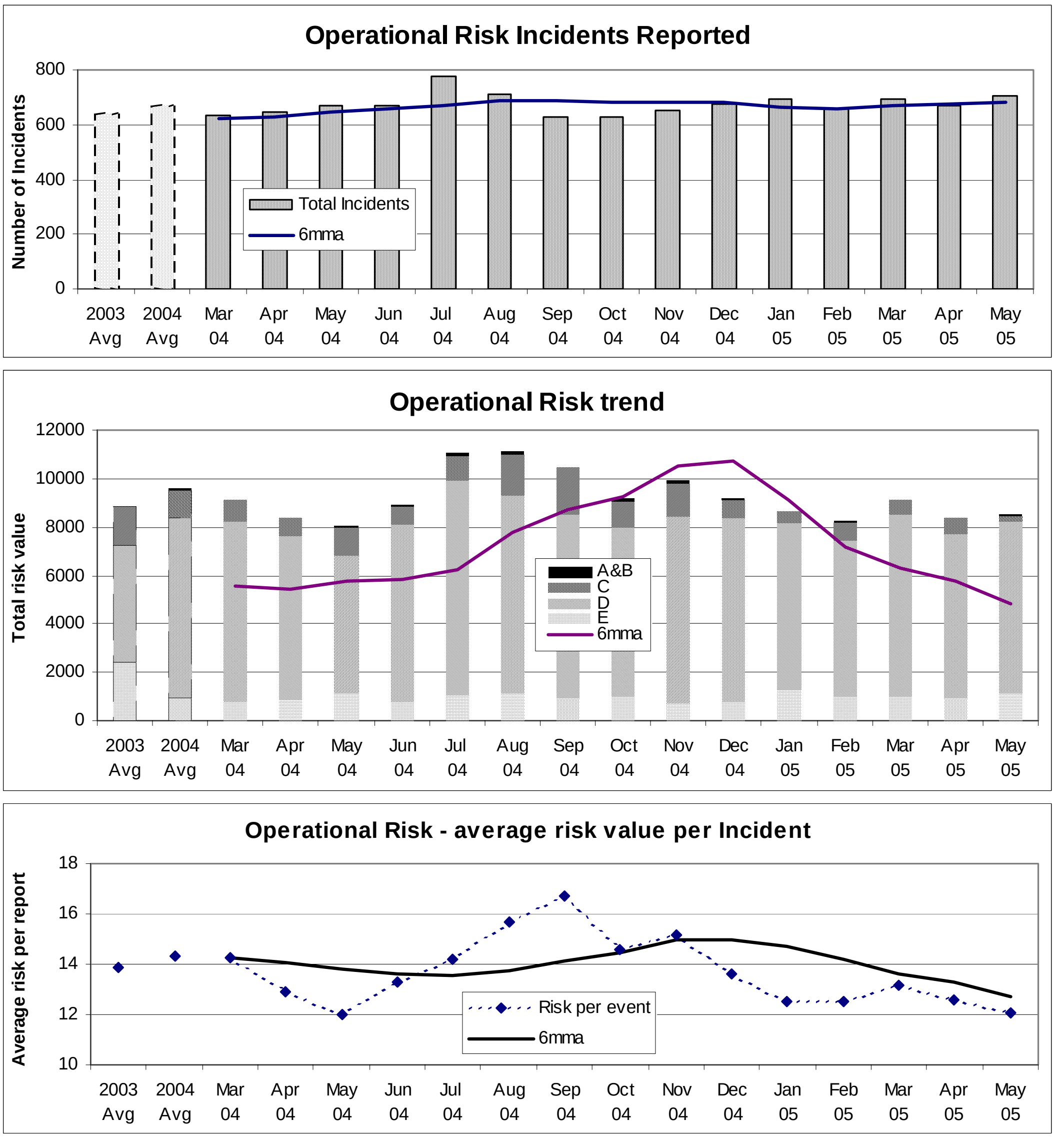

Figure 3 – Risk Trends

Figure 3 shows a way to present the overall risk trend to ensure that reporting rate is an integral part of any analysis of the data. The first graph shows the reports raised each month regarding operational safety with a six month moving average trend line. The second graph shows the cumulative risk value of all those reports divided into the A to E risk categories that they have been assigned. Each report is weighted depending upon its risk category so a lesser number of higher risk events are equivalent to a larger number of low risk events. The third graph shows the average risk value per report with a six month moving average trend.

From the first graph it can be seen that the reporting rate is fairly level over time, it may also be appropriate to normalise this data against flying hours to ensure that significant operational changes do not sway it. From the second graph it can be seen that the measured level of operational risk has reduced. The third graph validates this showing that the average risk per report is reducing which is a good indicator of both risk and reporting. The theory behind it being that the high risk, or at least the high severity, events cannot be hidden, therefore any marked reduction in reporting will manifest itself in an increase in risk per report. What it does not clearly control is where an increase of the reporting rate is driven by a change in reporting requirements, technology or a campaign of reporting of low risk type events. Such a change would need to be controlled by a review of the reporting environment and analysis of the events being reported.

Additionally the overall measure of risk is then based on the risk ratings given to individual events, a subjective activity that will vary widely between different people and parts of the organisation. Also this activity could be subject to external pressure as the lower the ratings the lower the overall ‘perceived’, and reported, risk. To overcome this problem requires both careful oversight and verification of the risk rating process, or its centralisation within the organisation.

The oversight process could range from a regular review of a sample of risk rated events by an experienced body of people, regular comparison of the risk rating between similar events over time, or the complete checking of the rating of all events. The alternative is to place the rating process under the control of a central body that has no direct interest in the resultant measures and that is remote from the operational areas. The problem with the second approach is that the skill and experience of the operational areas are vital in ensuring that the risks are properly recognised and appropriately rated. Ultimately the solution may be to centrally ‘own’ the process but to use the operational areas, with their skill and experience, to fulfil the task.

Using Risk Measures to focus resources

The best use of the data has to be to allow the organisation to focus attention and effort on the parts of the operation that indicate the most risk. With some of the verification efforts suggested earlier, the data should be reliable enough to be used in this way. With the limitations in the data you could potentially end up ‘looking down the wrong hole’ but at least you have used your safety data to best means and overall will almost certainly have looked down more of the right holes than the wrong ones.

To be able to divide up the risk by area requires each event to be appropriately classified for cause and other factors. This classification requires the selection of the type of event from a set of classifiers or descriptors that explain what the event was and what its likely cause was. An event may involve several causes and areas of the operation so each event may have several classifications. The analysis then requires each classifier to be assigned an 'owning' area of the operation so that the total risk can be divided between them with each event carrying its weighting of risk as assigned to it.

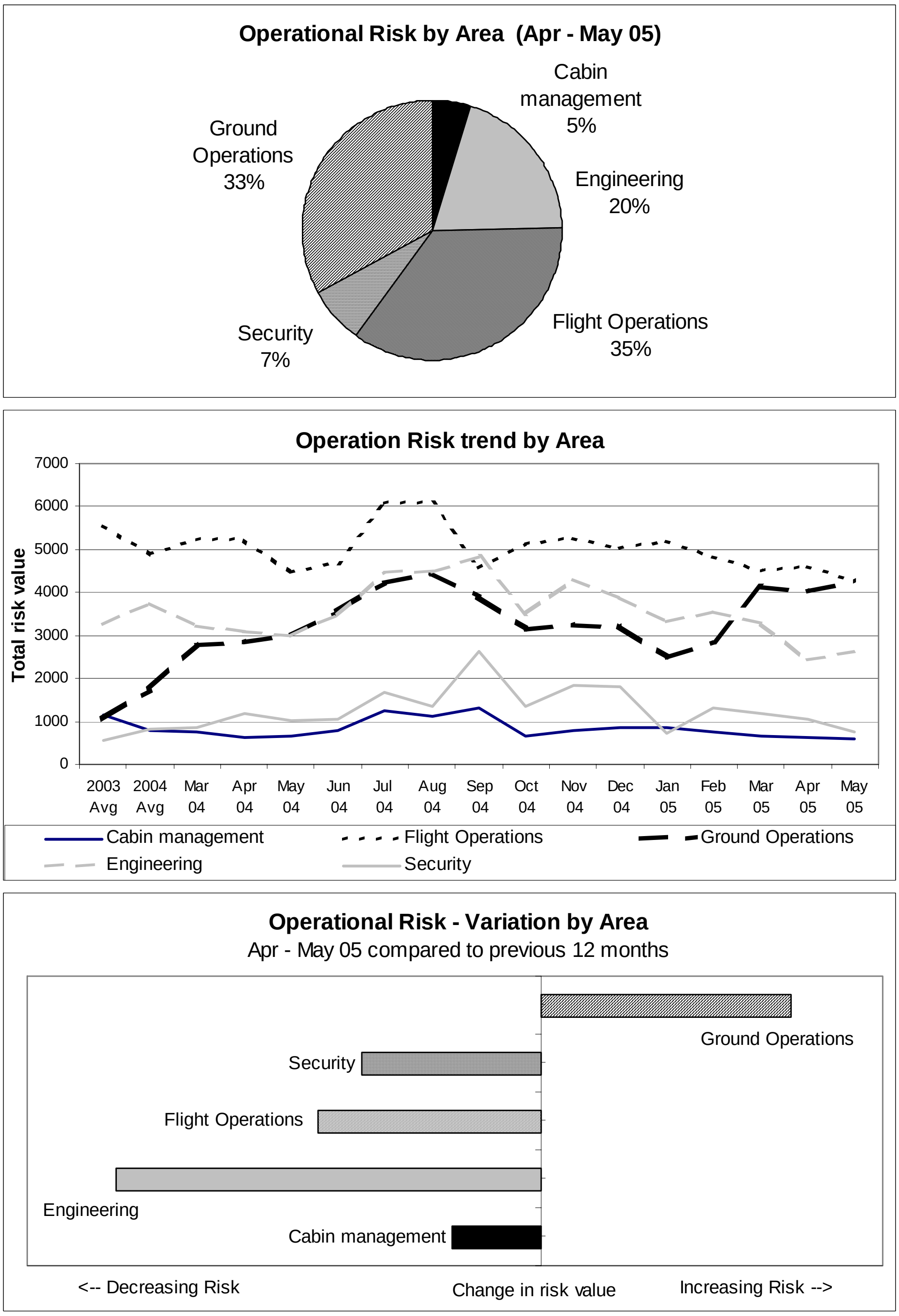

Figure 4 – Risk by Operational Area

Figure 4 shows a method of presenting the risk data to see the different levels of risk associated with various operational areas. The first graph gives an indication of the proportion of the total operational risk that is attributable to each area over the preceding two months. The second graph gives a trend of the risk associated with each of these areas and shows the levels or risk relative to each other. The final graph gives an overview of the shift of risk for each area by comparing the last two months with the preceding twelve months.

This presentation provides an overview of the distribution and shift of risk within the operation. It does not seek to define exactly where to look, as the data is not certain enough to support such assertions, but it does offer a number of indicators to allow the operational management to focus safety efforts in the right areas.

Once the data is structured to support such analysis it enables a similar analysis to be performed on the data at a lower level. For example each area of the risk pie chart can be further analysed to look at how the total risk for that area is distributed between types of events or particular causes. In the case of Ground Operations the analysis could encompass items such as Ramp Management, Hold Loading, Passenger Handling, Aircraft Movements. Each area can be analysed to allow the operational management to focus on individual areas of risk to enable the overall risk to the operation to be addressed effectively.

Conclusion

In a data driven operation it is clear that operational safety indicators of some form are of benefit to ensure that safety stays at the forefront of operational decisions. It is however evident that the data available to provide such metrics is partial and subject to many limitations.

Risk is a widely accepted classifier that is associated with safety data and as such provides a vehicle for providing an overall measure of risk of the operation. The most widely available source of such risk data is from incident reporting within the organisation but such reporting is subject to many pressures and variations. These pressures and variations mean that at best the data is only a partial view of the safety of the operation, but with careful management and monitoring such data could be used as a reasonable sample of the risks.

The presentation of this risk data in the form of trends needs to be done in a way that ensures that the wrong pressures are not applied to reporting and risk assessment. This requires the reporting rate to be an integral part of the measures and for the rating to be managed or monitored by a limited body of people.

Ultimately the greatest value in collecting safety data is to use it to reduce the likelihood of an accident or serious incident and therefore any measures derived from it should have this objective. The measures should compel examination of the areas of the operation with the most risk and then drive tangible actions to reduce that risk. The same measures will then allow you to monitor the effectiveness of those actions by the subsequent reduction in risk.

Acknowledgements

The author wishes to thank many Safety professionals at British Airways for their invaluable input into the work that inspired this paper.

Reference

1. Vick, S., *Degrees of belief: subjective probability and engineering judgment,* 2002.

2. Macrae, C., Pidgeon, N., and O'Leary, M., *Assessing the risk of flight safety incident reports.* Proceedings of a Workshop on the Investigation and Reporting of Incidents and Accidents: 99-106. University of Glasgow. 2002.

3. Rose, A., *'Free Lessons' in Aviation Safety,* Aircraft Engineering and Aerospace Technology, Vol.6, No.5, 2004.

4. Smithson, M., *Ignorance and Disasters.* International Journal of Mass Emergencies and Disasters, 8(3): 207-235, 1990.

Author Biographical Notes

Andrew Rose is a Chartered Engineer with the Institute of Measurement and Control and a member of the institutes Aviation Panel. Andrew has worked for British Airways for 19 years, starting as an Engineering apprentice before gaining a Masters Degree in Control, Instrumentation and Systems Engineering from City University, London. Andrew now works as part time as an Air Safety Investigator for British Airways and also runs Llanbury Consulting, an aviation consultancy business.

From 1992 to 2001 Andrew worked in various roles within the Avionic Standards and Design areas, where he was responsible for Air Traffic Surveillance and Airborne Collision Avoidance Systems (ACAS). Whilst in the Standards Department Andrew held UK CAA delegated Design Authority for the approval of aircraft modifications and for a number of years chaired the Airlines Electronic Engineering Committee - Airborne Separation Assurance Systems Subcommittee.

Since joining the Airline Safety Department in 2001, Andrew has undertaken a number of investigations into significant incidents and has taken a major role in analysing data and developing safety metrics.