CHALLENGE OF SAFETY DATA ANALYSIS – TOP MODELS WANTED

or "Don't call me a cab, when I ask for a map"

Jari Nisula

Operational Monitoring & Human Factors, Flight Operations Support, Airbus[16]

Abstract: Modern Flight Operations Monitoring tools have satisfied the hunger for safety data on minor events and deviations – and thus set the scene for very proactive safety management. There remains, however, the conceptual challenge of how to learn effectively from such data. While the initial case-by-case analysis is usually straightforward, the classical keyword-based analysis methods only give limited support to a more proactive trending type of analysis on the whole database of events. The paper suggests that new methods based on top-down safety models are a very promising option in facing the challenge. The promises and main uncertainties of such methods are discussed. The paper also argues that the term "risk management" is often used lightly in the context of safety processes, which are far from rigorous and systematic risk assessment and management.

Key words: Incident/accident analysis, aviation

1. INTRODUCTION

The aviation industry has traditionally used safety related data from everyday operations to make the aviation system safer. Reporting, data analysis and corrective actions are carried out by different organizations and at different levels.

[16] The views expressed in this paper are solely the personal opinions of the author and do not bind or necessarily reflect those of Airbus, any of its affiliates, or its advisors.

The last decade, and especially the last few years have gradually changed the methods, tools and available data to a significant degree. Some traditional difficulties were washed away almost in one go, just to show the underlying challenges clearer than ever.

It has been clear for a long time that accidents and serious incidents are too rare and random to be used as the single source for safety analysis and lessons learned. Different kinds of reporting systems have been used for decades to facilitate identification of threats before they escalate to serious incidents. The recent arrival of powerful Flight Operations Monitoring (FOM) tools has suddenly enabled the collection of vast amounts of detailed information covering virtually every single flight of an airline. Flight Data Monitoring (FDM) is a process where hundreds of flight parameters are constantly recorded on Quick Access Recorders (QAR) where the data is stored on optical discs. The discs are read by dedicated software at the airline, and pre-defined events are automatically detected. Specialists then carry out further analysis. FDM has already been mandated in some countries, and year 2005 will see a big part of the world mandating operators to run a FDM program. FDM has been complemented with observation techniques like LOSA (Line Operations Safety Audit) or the Airbus-developed LOAS (Line Operations Assessment System), which are based on human-made observations on the flight deck. Simultaneously, software tools have arrived to improve traditional reporting systems.

These new techniques and methods have established an extremely rich source of safety data, and – quite importantly - pushed the focus towards more and more minor events, like minor deviations from ideal flight path during final approach or excessive bank angles at lower altitudes. This extension of focus has underlined the challenges of analyzing data where the content is no longer a long sequence of events with a known (harmful) outcome, but rather a single (causal) factor, which in itself is not damaging, but could be in another context. How do we assess the risk of such an event? How do we store event information so that it can be used effectively in later analyses? The bottom line is: how do we prioritize safety actions. These challenges form the guiding theme of this paper.

The terms "Risk Management" or "Safety Management" are gradually becoming the reference for managing flight safety activities in an airline (Civil Aviation Authority 2002, Workshop on Risk Analysis and Safety 2002). Ideally, risk management would require taking a wide and holistic look in all safety threats for the activity in question, assessing different risks, and then tackling them – starting from the issue associated with the highest risk. The growing demand for such an approach highlights the difficulty in

trying to allocate risk to minor "everyday" events and threats. The situation is not made easier by the key terms themselves: "risk" and "safety" are abstract terms with much more in-built complications than their casual everyday use suggests (Nisula 2002). People talk fashionably about "risk management" when a detailed look into the concept would reveal just how difficult true risk management would be.

Despite the huge power of modern FOM tools and some very successful safety programs – especially FDM related – one could argue that the overall efficiency ratio of safety data analysis is not impressively high (Paries et al. 1996). Vast amounts of data are reported, collected, processed and analyzed without ever resulting in any improvement actions. On the other hand, numerous chances to learn more are certainly missed because patterns are not recognized in the vast databases of safety data – or because some valuable aspects of the events were not recorded in the first place.

This paper argues that the barrier of inefficiency is not that of lacking technical means, but that of lacking conceptual models. We need conceptual tools to show how different functions contribute to safe flight operations – and how different elements of threat and risk can combine to create incidents and accidents. Only such models can help us learn efficiently also from minor events. We do not need yet another piece of data, we need a map to show us where that piece of data plays a role.

2. TYPICAL ANALYSIS METHODS AND THE CHALLENGE

Whatever the source of the safety data, one can usually distinguish two[17] levels of analysis performed on them: case-by-case analysis (or clinical analysis) and the long-term analysis.

[17] We could consider the analysis of Flight Data to represent a third type of analysis. Significant events are analyzed one by one, but the main analysis is based on statistics on different event types, their trends and correlations with other event types. The statistics lead the analyst to take a deeper look into some particular issues and even in some individual flights – the process is thus a constant switch between statistical and clinical analyses. Regrettably, the scope of this paper does not allow a detailed discussion of this quite successful analysis process. See [9] for a discussion of the different aspects of the FDM process.

2.1 Case-by-case analysis

Typically, a flight safety manager receives an Air Safety Report (ASR) and reads it through. S/he sees if there are elements requiring immediate actions and tries to identify key issues in the report and perhaps allocate related keywords before storing the report in an ASR database. If further investigation is needed, s/he will coordinate getting the necessary inputs from all applicable departments.

Similarly, within an aerospace manufacturer, incoming reports are often analyzed in expert teams, ensuring the coverage of all aspects of the events (design, training, procedures, pilot proficiency, human factors, system knowledge, etc.).

The case-by-case analysis relies mainly on the technical (and human factor) skills of the people involved, and one could say that with a proper team with good knowledge, it is probably fulfilling the function well. Safety issues are identified and actioned.

There are two main reasons why limiting safety work to case-by-case analysis is insufficient by itself. First, treating every issue as a single case does not allow the safety manager to see the "big picture": the underlying patterns and problems and the development of different issues over time. Secondly, the scope of the events analyzed this way is often limited to the more serious events - an expert team is not called to analyze every minor deviation detected in the flight data. Consequently, this process is not proactive enough to be used as the only safety process. In line with this, a major accident investigation can be seen as a large-scale case-by-case analysis of one event.

2.2 Long-term analysis

The real challenge is the analysis based on a larger set of safety data – data where each event[18] has usually already been analyzed in a case-by-case manner when it was entered into the database. The long-term analysis should unfold hidden patterns in the safety data and create lessons learned which were not derived from the case-by-case analysis. Experience shows that this is not an easy job to do. Figures from aviation and other similar safety-minded industries (e.g. nuclear) show that typically less than 5 % of data in

[18] The term "event" is used here because most databases deal only with events – unfortunately. We could as well talk about "safety issues" or "safety concerns".

safety databases are used for launching concrete actions based on the long-term analysis (Safety data analysis workshop 2003).

The term *long-term analysis* is used here for two reasons. First, this type of analysis requires a reasonable amount of data - usually obtained over a long time period. Secondly, the term *statistical analysis* was deliberately avoided because its scope is too restricted. While statistics on event facts (e.g. weather, destination, time) and on some keywords can be invaluable, they are not enough to respond to the full challenge of long-term analysis. In the complex flight operations environment, the prioritization of safety issues cannot be done simply by comparing their frequencies. Similarly, even a clearly increasing frequency of a certain event type is not necessarily enough alone to justify a safety action. This also explains the author's persisting allergy to the term "trend analysis".

There is a link between the type of safety events to be analyzed and these analysis methods. Higher-level events (e.g. incidents, accidents) give ample material for a rich case-by-case analysis, whereas the "bits and pieces" of very minor occurrences do not really lend themselves for this analysis method. As said earlier, the safety management focus has been moving more and more towards these "bits and pieces" of safety information, and the pressure to find suitable analysis methods is increasing.

2.3 The challenge

Let's look at the core of the challenge: the type of safety data that a safety manager at an airline is typically dealing with.

Example 1: *"An 80-year old passenger found smoking in the toilet"*

Example 2: *"In cruise at flight level 350 the aircraft encountered standing wave activity and moderate turbulence. A cabin crew member was hurt while taking her seat when ordered to do so."*

Example 3: *"Crew: False Localizer capture ILS*[19] *15 following [approach procedure] arrival, failure occurred just before [location]"*

Example 4: *"Unsecured cargo pallet in rear hold."*

[19] ILS = Instrument Landing System

The reader can appreciate the difficulty of risk assessment of such minor events by trying to decide which of the events carries the highest risk. Asking other people to repeat the exercise reveals how subjective the assessment is.

Each of the above events contains one element of risk (or "causal factor"). The element alone – on a good day – is not enough to cause an accident, and *did not* cause an accident in the real life event either. However, such factors might cause an accident *together* with some other *additional* elements and/or in another context. The questions to answer now are[20]:

- What are those other elements?
- What would be their probability to combine with the reported element?
- What is then the *risk* involved with these events?
- How can we store the core information in these reports in such a way that it can contribute to safety lessons in the future, e.g. through the long-term analysis?

Before moving further, we must examine the concept of risk. Engineers see risk as the product of the *probability* of an outcome and the *severity* of that outcome. When we try to assess "the risk" in a historical event, what are we actually doing? Factually speaking, as long as traveling through time is ruled out, the risk that a historical event (which did not end in an accident) would end in an accident is zero. What we really are after is: "what is the risk that *something similar* in the future ends in an accident?" Using example 1, we would ask: "what is the risk that in the future a passenger will smoke in a toilet and the sequence will end in an accident?"

Answering such a question forces us to build imaginary scenarios, which develop into accidents thanks to additional risk elements (or "threats" or "causal factors"). The probability and the severity of each scenario outcome are a function of our choice of additional elements. To continue with the previous example, we can build different scenarios:

[20] The situation is quite the opposite in the clinical analysis of accidents (and the like) because the event sequence ties together all causal elements, disclosing their roles and consequences. This creates the opposite problem that in hindsight it seems like the sequence "couldn't have gone any other way", i.e. the probability of other causal paths and outcomes is underestimated – leading to the unjustified conclusion of "they should have seen this coming". This phenomenon is called the Hindsight Bias [11].

Scenario 1: *Passenger smoking in the toilet, detected by the smoke detection unit and/or the cabin crew. No consequences.*

Scenario 2: *Passenger smoking in the toilet during approach, detected by cabin crew but delaying the cabin preparation and distracting the cabin crew, thus leaving some bottles and glasses in the cabin, and some passengers fastening their seat belts late during final approach, creating the potential for injuries in case of turbulence.*

Scenario 3: *Passenger smoking in the toilet during cruise flight over the Pacific, paper towels in toilet catching fire, poor crew performance leading to spreading of the fire to the cabin, resulting in several fatalities due to smoke or loss of control of the aircraft.*

Using our imagination this way, we can see that *virtually any event* allows us to move from a "probable-non severe" scenario to a "very improbable-very severe" scenario, just by adding elements, i.e. varying what we include in *"something similar"*. The resulting scenarios also have different *risk* levels.

How do people then allocate risk to such events in real life? Many software tools offer some type of probability-severity matrix for this purpose, but how do analysts pick the "right" square? They must decide how remote scenarios they still consider reasonably possible, and then estimate their probability and the severity of the related outcome. While doing so, they are faced with several problems. Building scenarios, estimating probabilities[21] and judging severities are all based on the analyst's subjective, implicit and partly unconscious models of safety in the system under study. Biases and heuristics further distort the estimates. Recent discussions that the author had with some flight safety managers reveal that even their basic strategies for scenario building and risk assessment differ fundamentally: some want to stick to the event exactly as it happened in real life without any further scenario-building, whereas others stress they want to think about what *"could have,* realistically, happened". Here the word *realistically* is loaded with all the above-mentioned subjectivity. Furthermore, the analysts would only cover scenarios that they *know about,* and *think of* at the time of the analysis. They would also be unable to explore the different scenarios with their probability-severity combinations

[21] It frequently happens that analysts have to revisit a probability estimate because new events prove the scenario more probable than expected.

systematically like a computer – the human result is more at the level of a *feeling*.

Based on these limitations, the author argues that the validity and value of applying the current simplistic approaches in the risk assessment of minor events are questionable.

What is missing is a framework, which shows the roles of causal factors in different accident scenarios. We must acknowledge that for risk assessment, analysts always refer to "safety models". It is our responsibility to replace the subjective, implicit models by shared, explicit models, which can be built and challenged by expert groups. The real challenge is thus on the conceptual side of creating meaningful and practical safety models.

3. LOOKING FOR SOLUTIONS

The traditional answer for organizing the chaos of safety data has been the use of keywords. Typically, database entries are characterized with keywords, in the hope that these would catch the essence of what happened – and in the hope they would facilitate the long-term analysis by keyword-filtering. This approach has several known drawbacks:

- **Subjectivity:** Use of keywords and risk classification is not consistent between different analysts, even when trained identically.
- **Novelty of results:** Any keyword set is a self-fulfilling prophecy, because non-existing keywords cannot be used and matters represented as keywords may attract extra attention.
- **Causal synergy:** The living dynamism of the event sequence is lost. The database "does not care" if two keywords come from the same event or from two different events.
- **Causality:** It is not always clear which factors can be accepted as causes of an event and where to stop in the exploration of the causal sequence ("the big-bang syndrome").
- **Capturing positive lessons:** Keywords usually reflect negative aspects, failures. The valuable lessons of positive aspects and successes are largely missed.
- **Risk management:** A keyword structure typically does not provide any framework for risk assessment.
- **Productivity:** Typically only 5% of data is re-used thanks to the keywords (Safety data analysis workshop 2003). Most databases

also have a significant percentage of events without any keyword allocation.

Ambitious keyword systems easily become unpractical monsters. In the late 1990's The Australian Bureau of Air Safety Investigation (then called BASI, now part of the ATSB[22]) was studying a new safety reporting and analysis tool SIAM[23]. The old tool OASIS[24] had an impressive keyword structure of 1400 descriptive factors. The study revealed the following shocking but not unusual facts (Lee & De Landre 2000):

- On an average year, only 50% of descriptors were used
- 29% of descriptors were not used at all
- If 75% of least used descriptors had been removed, it would have affected only 0.5% of the events!!

Either we have to accept the highly imperfect situation where the main focus is on the case-by-case analysis and risk assessment is based more on personal intuition than a robust method, or we have to find completely new approaches to safety data analysis.

4. TOP-DOWN APPROACHES IN SAFETY MANAGEMENT

Perhaps the most promising way to go is the development of so-called top-down analysis methods. The idea is that before starting to analyse event data, models are created, showing how different factors contribute to different accident types. Typically, this is achieved by making a structured presentation of the vital safety functions necessary for safe operation (for the scenario in question). There is a finite number of vital safety functions, whereas trying to directly model the failure conditions would be an endless task. Safety functions can be developed into more detailed structures of safety assumptions (or *safety principles*) with the help of well-known techniques like FMEA[25]. Safety principles are positive statements like "technical quality of radio communication is acceptable", The model, which represents the "a priori" understanding of the situation, is then put to the test of reality by feeding it with safety data. Safety lessons come through changes in the model – either in its structure or in the confidence given to

[22] Australian Transport Safety Bureau
[23] Systemic Incident Analysis Model
[24] Occurrence Analysis and Safety Information System
[25] Failure Mode and Effect Analysis

particular safety principles. An advantage of the positive statements is that one can easily capture both negative *and positive* lessons from safety data. The models also highlight the high number of assumptions made on the human operators – assumptions, which are often not challenged enough.

This approach differs fundamentally from classical bottom-up approaches, where the raw data itself is used to help define the classes or keywords, which are then used to organize the data.

4.1 Promises of the top-down approach

The top-down approach is promising, because it could overcome most drawbacks of classical bottom-up applications:

- **Subjectivity:** A common, expert-group-made safety model is more accurate than individual, subjective models
- **Novelty of results:** The analysis process is opening questions, rather than trying to match the event with predefined keywords
- **Causal synergy:** The model is not used to re-create the events. The full event narratives are used.
- **Causality:** The safety model presents the *common and official defenses and safety assumptions,* thus any factor handicapping or eliminating them *is* a causal factor, and factors not presented on the model *are not* causal factors[26].
- **Capturing positive lessons:** Success of safety principles in each event are also recorded
- **Risk management:** The place of each safety principle in the overall safety model is visible, which gives a good starting point for risk management.
- **Productivity:** Even minor events typically link with at least one safety assumption, and lessons are captured from virtually all pieces of safety data.

[26] For example, the skill requirement of a typical line pilot to perform an ILS approach would be represented on the model, and a failure to perform an ILS approach would be considered a causal factor. Things like improvised flying techniques which could have prevented a certain accident would not be listed on the model, because they are not standard requirements for pilots, and the failure to perform such an in-hindsight-obvious escape maneuver would not be accepted as a causal factor.

4.2 Examples of top-down safety data analysis methods

There have been several attempts to apply top-down safety models – either using numerical probability values or more qualitative measures[27].

One of the first such tools in civil aviation was the British Airways–developed RATBAG (Risk Analysis Tool for British Airways Group). While looking like a classical Probabilistic Risk Assessment tool, it must be given the credit of having produced a large top-down framework of positive safety principles. The nominal outputs of RATBAG are updated probability figures. This enables the user to simulate the impact of changes in the operation. However, just having the qualitative model itself is at least half of the benefit (Savage 2000). Figure 1 illustrates the RATBAG model.

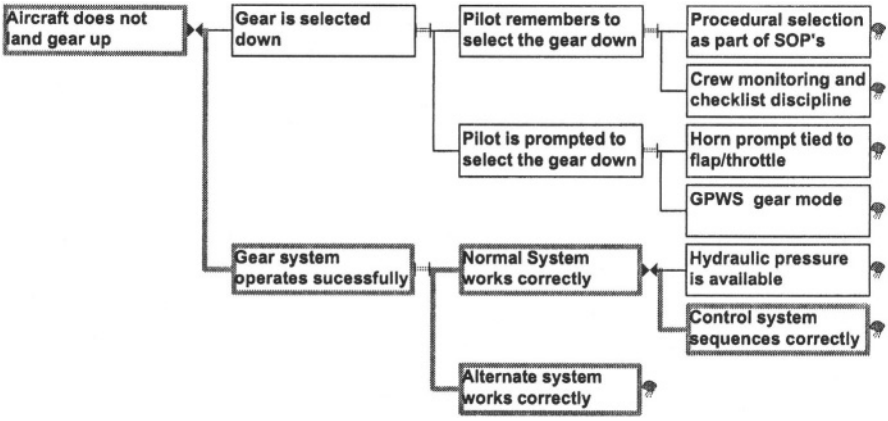

Figure 1. Extract of a RATBAG model (Savage 2000).

Perhaps the most advanced and detailed work on this domain has been carried out by Airbus, Eurocontrol and the ADAMS[28] 2 research team, in three separate projects – all supported by the consulting company Dedale SA. A FMEA process was applied on high-level safety functions for a safe flight to identify initiating situations to accidents ("threat" in figure 2). A more precise content was defined for each initiator for each flight phase. The full model would consist of one Safety Architecture (SA) (see figure 2) per initiator per flight phase. Each SA contains a structure of safety principles (SP) whose role is to first prevent reaching the initiator (prevention), then

[27] It is interesting to note that the analysis power of Flight Data Monitoring is partly based on the fact that FDM is in some aspects a top-down method. In FDM, deviations are measured against pre-defined reference flight profiles: positive definitions of safe performance. This also forms the basis for risk assessment in FDM.

[28] Aircraft Dispatch and Maintenance Safety; under the European Commission 5th framework

prevent reaching the accident (recovery), and finally, in the case of an accident, mitigate its consequences (mitigation). Safety principles can be broken down to sub-structures of more detailed safety principles. For example, an SP "thrust asymmetry is detected" could split into two sub-SP's covering the detection by the pilots and by the aircraft systems. The Safety Architectures cover all aspects coming into play in the scenarios: aircraft, human operators and the environment. This sets high demands to the expert group responsible for building them.

Figure 2. A Safety Architecture gives a structured presentation of the Safety Principles protecting the aircraft against a given threat and accident type, in a given flight phase.

The ultimate idea would be to have a software tool incorporating all Safety Architectures and all necessary functions to implement the methodology. The analyst would be guided through a single event analysis with the help of the corresponding Safety Architecture(s). The tool would keep track of how many times different safety principles have failed and gradually build a global picture of all accident scenarios and the robustness of the corresponding defenses (SP's).

The initial experience of event analysis with the method was promising, but evaluating safety decision-making using the method would require a large set of Safety Architectures, which nobody has for the time being. The full set of SA's would amount to about 200 and represent a demanding and expensive exercise requiring long-term allocation of many experts.

4.3 The future

Even with the considerable effort put in the development of top-down safety analysis methods there are still several uncertainties about the final usefulness of such methods. Can such models represent the complexity of real life to the extent that the analysis results are reasonably correct? What

are the resources needed to create the initial a priori safety models, keep them up to date and do event analysis? Can this method detect issues, which would otherwise be missed? Many people are optimistic, but the final answers could only be given with a functional tool, which is man-years away. There is also the challenge of breaking existing mentalities: it is not easy to start thinking in terms of positive safety principles after years of thinking in terms of failures.

As the full top-down model for aviation does not seem to be available very soon, the development of local models should be encouraged. Many organizations have developed models for their own limited needs. The ADAMS 2 program applied the approach on a few maintenance error scenarios, and Eurocontrol on level busts and ATM[29]-related safety devices. Some airlines have identified precursors to different accident types and the Flight Safety Foundation has hosted the development of the Flight Operations Risk Assessment System (FORAS) (Hadji et al. 2002), which has looked into CFIT[30] scenarios. Such local models could one day contribute to a more global top-down safety model – especially if the developers keep this option in mind when developing their local models.

Concerning risk assessment of minor events in the absence of a top-down model, one could hope that we could move away from the simplistic method where the analyst picks one square from the risk matrix to represent the event. We could try to develop a method where the analyst studies systematically several additional causal factors and the resulting scenarios. Each scenario would be rated in terms of probability and severity, and the risk of the original event would be derived as a combination of these probabilities. Existing data in the database could be used to help estimate the probabilities and severities.

5. CONCLUSIONS

Terms like safety and risk, which are used very casually in everyday conversations, are in fact difficult concepts to work with. Many current practices of "risk management" would not fulfill the criteria of scientific rigor: too often risk assessment is based on rough subjective estimates.

[29] Air Traffic Management

[30] Controlled Flight Into Terrain: an airworthy aircraft under the control of the flight crew is flown unintentionally into terrain, obstacles or water, usually with no prior awareness by the crew.

The challenge of learning from safety data and *managing safety* is especially big when trying to deal with long-term analysis of multiple minor events, i.e. trying to learn extra lessons from a database of minor safety issues, which have already been analyzed clinically. Classical methods – often based on keywords – are quite disappointing in this sense.

Based on the initial experience, the so-called top-down methods seem promising in many ways, but practical applications of these methods are not visible in the near future. The final question to be answered is whether such methods optimize the effort invested in risk management. Any improvements in modeling accident scenarios and structuring risk assessment should be welcomed. Hopefully, the switch to next generation methods in both areas comes soon.

REFERENCES

Civil Aviation Authority (2002) (UK) – Safety Regulation Group. *CAP 712 – Safety Management Systems for Commercial Air Transport Operations.* Civil Aviation Authority. West Sussex, UK, 2002.

Hadji M. et al.(2002) Flight Operations Risk Assessment System (FORAS). *Proceedings of the Joint meeting of the FSF 55th Annual International Air Safety Seminar IASS, IFA 32nd International Conference, and IATA, Dublin, November 4-7.* Pages 367-373.

Lee R. & Joanne De Landre. (2000) *Systemic Incident Analysis Model (SIAM) – A new approach to safety information.* BASI (now ATSB). Presentation.

Nisula J. (2001) *Efficient use of Flight Data Monitoring as a part of airline Flight Safety Management.* Airbus, Blagnac, France, 2001. (copies available from jari.nisula@airbus.com)

Nisula J. (2002) *Flight Safety Management: Towards Risk Assessment and Safety Measurement.* Presentation given to the 15th Airbus Human Factors Symposium, Dubai, 18-20 June 2002. Airbus, Blagnac, France, 2002. (Symposium CD-ROM's available from jari.nisula@airbus.com)

Paries J. et al. (1996). *Development of a Methodology for Operational Incident Reporting and Analysis Systems; Final Report.* Direction Générale de l'Aviation Civile (DGAC). 1996. Pages 5-9.

Proceedings of the fourth Workshop on Risk Analysis and Safety Performance Measurement in Aviation. Atlantic City, New Jersey, August 27-29, 2002.

Proceedings of the National Workshop on Risk Analysis and Safety Performance Measurement in Commercial Air Transportation. Rutgers University, New Jersey, July 20-22, 1999.

Proceedings of the second Workshop on Risk Analysis and Safety Performance Measurement in Aviation. FAA William J. Hughes Technical Center, Atlantic City, New Jersey, August 22-23, 2000.

Proceedings of the third Workshop on Risk Analysis and Safety Performance Measurement in Aviation. 2001

Safety data analysis workshop at CENA 2003) (Centre d'Etudes de la Navigation Aérienne), Toulouse, 11.3.2003.

Savage J. (2000). Risk Analysis by Dependency Modelling. Presentation to the (British Airways Safety Information System) *BASIS User Conference 2000.* British Airways.

Transport Canada (2001). *Introduction to Safety Management Systems.* TP 13739E, 04.2001. Transport Canada, Ottawa, Canada.

Woods et al.(1994) *Behind Human Error: Cognitive systems, Computers and Hindsight.* CSERIAC State-of-the-art Report 94-01. The Ohio State University. Pages 177-183.