2009

Risk Assessment in Aviation

Robert Mauro

Immanuel Barshi

Risk assessment in aviation

Robert Mauro
Decision Research &University of Oregon
Eugene, Oregon, USA
Immanuel Barshi
NASA Ames Research Center
Moffett Field, California, USA

In aviation, many actions are taken to reduce risk. However, not all risks can be avoided. To effectively manage risk, managers and regulators must evaluate and compare risks associated with different threats. Yet, it is frequently difficult to obtain reasonable assessments of these risks. Traditional approaches often produce unsatisfactory results when the probability of failure is low but the costs of failure are high -- as is often the case in modern civil aviation. Attempts to use a single dimension to evaluate threats often lead to unreliable and contentious assessments. Many risk assessment heuristics and displays can yield misleading and sometimes mathematically incongruous assessments. Furthermore, increases in costs caused by people's reactions to failures are often ignored or grossly underestimated. In this paper, problems with risk assessment in aviation are discussed and a Tool for Risk Identification, Assessment, and Display (TRIAD) designed to address many of these problems is described.

In aviation, safety and efficiency are primary goals. Many of the actions taken by aviation professionals are taken to reduce risk. However, one cannot avoid all risk. Regulators and managers must frequently decide which potential problems to address. To effectively manage risk, one must be able to evaluate the risks associated with different threats and compare them. But it is frequently difficult to obtain precise assessments. To accurately assess the risk associated with a potential failure or other threat, one must consider the possible outcomes that could occur, the likelihood of each outcome, and the consequences that may be associated with each outcome. In this paper, we discuss the assessment of each of these aspects of risk and describe a Tool for Risk Identification, Assessment, and Display (TRIAD) that was designed to assist in their assessment.

Risk is generally defined as a combination of likelihood and consequences -- the more damage that may occur, the greater the risk; the more likely a threat, the greater the risk. In traditional probabilistic risk assessment (PRA), risk is quantified by multiplying an estimate of the amount of potential damage by the estimated probability of the threat (e.g., Bier & Cox, 2007). In many cases, this assessment can be accomplished simply and the obtained result matches our intuitions. For example, a computer manufacturer may be able to estimate the probability that a microchip will fail within a warranty period quite precisely based on laboratory and field data. Calculating the cost of a new chip and the labor required to replace it is also relatively straightforward. Hence, the risk posed to the manufacturer by the potential failure of the microchip can be easily assessed. However, in many cases assessing risk is much more difficult.

Assessing the risk associated with a possible failure or other event becomes more difficult when:

- The event of interest (e.g., a failure) can have many possible outcomes.
- The event under consideration is not repeatable or there is no data from which to directly estimate the probability of the event.
- The event could lead to different types of damage which cannot be easily measured on a common scale.

- The cost of the potential damage or the likelihood of the event is extreme. This is particularly noticeable when dealing with extremely unlikely events that could have catastrophic consequences.

Aviation typically operates under these conditions. For example, an airline may be concerned with a rash of pilot reports of anomalies in the operation of their new flight management systems (FMS). Given the financial and personnel demands of daily operations, management must decide how much time and money to invest in determining the cause(s) of these reports and finding a solution. This requires an assessment of the risk posed by the reported anomalies. In the worst case, an FMS problem could lead to a controlled flight into terrain (CFIT) accident. But there are many other possible outcomes. A CFIT accident is not likely unless the anomaly occurs on approach or shortly after departure. However, encountering an anomaly en route is not without cost. Most of the time, the pilots may notice the anomaly and correct it, but if they don't -- fuel will be wasted, the pilots and airline may be the target of FAA enforcement actions, and there is a (very low) risk of a midair collision. Furthermore, the anomaly could distract the pilots at an inopportune time and cause other problems. All of these possibilities must be considered.

Possible Outcomes

To accurately assess the risk posed by a potential problem, one must first consider the possible outcomes that could result if the problem were to occur. Often, individuals attempt to simplify this task by considering only the worst case. This can be misleading. For example, consider a hypothetical error in an airline's weight and balance calculation program. In the worst case, the aircraft could depart out of balance and encounter an event that causes the aircraft to enter a stall from which recovery is impossible given the weight distribution. However, this is an exceedingly unlikely scenario. A manager might reasonably conclude that this possibility is so remote, that other problems have a higher priority. However, there is a much more likely outcome that should catch the manager's attention. Aircraft that are flown "out of CG" may burn substantially more fuel because of the out-of-balance condition. Although this outcome is not catastrophic, over a large number of flights the cost of the error could be large enough to cause substantial financial damage to the airline. It is also not sufficient to consider only the most likely outcome. In many cases, unlikely outcomes have sufficiently serious consequences and are likely enough to be cause for concern.

Generating lists of possible outcomes requires domain knowledge and creativity. However, in many cases, one can generate outcomes by systematically considering the general classes of factors that are likely to affect the result of a failure or other problem. These factors include:

<u>Phase of flight</u> – The point during an operation at which a problem occurs can have substantial effects on the possible outcomes. For example, the failure of a critical component of a navigation system may have different consequences during takeoff/climb-out, en route, or during descent/landing.

<u>Time</u> – When a problem occurs can have substantial effects on the possible outcomes. For example, the failure of a component may have different consequences during the day, or during the night. Likewise, the same failure could have very different consequences for winter operations than for summer.

<u>Geography</u> – Where a problem occurs can affect the possible outcomes. For example, the failure of a critical component of a navigation system may have different consequences depending on whether the failure occurs over land or during a trans-oceanic flight.

Damage – The physical characteristics of the damage caused by a problem may affect the outcome. The result of a problem may be different if the physical characteristics of the damage (e.g., size, depth, location, and frequency) differ. For example, the damage caused by debris from a turbine engine failure may be different depending on the size and depth of the penetration.

Design Characteristics –The way in which a system is designed will affect the possible outcomes that could result from a problem. For example, the result of the failure of a given system may differ depending on whether the aircraft is equipped with a backup system. Likewise, consequences of a failure could be very different depending on whether the failure is annunciated to the crew or not.

Procedures and training – A problem can have very different outcomes depending on whether or not procedures exist for dealing with it, and on whether or not crews are trained to deal with it. (Note also that procedures and training are often used as interventions to reduce risk.)

Environmental Conditions – Environmental conditions, such as temperature, humidity, wind speed and direction, etc. can affect the result of a problem. For example, the effect of a failure in a cooling system may depend on whether the device is at or below a critical temperature when the system fails. Similarly, a failure in an ice detection system would have very different consequences if the flight is conducted in icing conditions, than if it is conducted in non-icing conditions.

Likelihood

To proceed with a risk assessment, one must estimate how likely it is that each possible outcome will occur. Sometimes, the probability of a given outcome can be estimated quite precisely. For example, one may have engineering data that indicate how often a component fails in practice. But often this is not the case. Many likelihood assessments must be based on expert judgments. In many cases, experts will be reluctant or unable to specify a precise probability for a possible outcome. For example, an engineer may be able to specify the conditions under which a component of a navigation system will fail but no one may know how often those conditions occur in practice. However, even in these instances, it is rarely the case that one knows nothing. It is rarely the case that the probability of an outcome could *plausibly* range from zero to one. Even when one cannot estimate the probability associated with an outcome precisely, one can often offer a "best estimate" and specify a range around that estimate that will confidently bracket the actual probability. This is sufficient to continue with the risk assessment.

Consequences

Because risk is a function of likelihood and consequence, the possible damage that could result from an event must be assessed. In the microchip example used above, it was relatively easy to assess the possible damage because the costs are easy to calculate and only one type of damage, monetary loss, was considered. However, an event could cause many different types of damage that are not easily measured on a single scale. An event could cause property damage, injury or loss of life, or disrupt operations. Furthermore, an event could generate secondary damage through people's reactions to the original event.

People often attempt to simplify the assessment process by trying to use one measure to scale all of the different types of damage. For example, insurance companies and international agreements specify how much the loss of a limb or the death of an airline passenger is worth in dollars. These amounts can then be combined together with estimates of the costs of property damage and lost revenues to arrive at a single monetary value that can be used as *the* measure of the consequences of an accident. However, attempting to create a single scale on which all potential consequences can be arrayed may be counter-productive. For example, people may reasonably disagree with the value attached to life by an insurance company; courts often do. Furthermore, these calculation may lead decision-makers to make trade-offs that they themselves find unacceptable. For example, if an arbitrary monetary value is attached to the value of a life, then the rational decision is to forgo safety investments whenever the costs of those investments exceeds the monetary value of the lives likely to be lost if the investment is not made. Once a monetary value for a life is accepted, the trade-off appears rational although the individuals making the decision may not agree that the value of a life can be reduced to the specified amount.

Disagreements about the validity of an assessment may arise not because of any debate over the possible consequences or their likelihood but only over the value attached to the consequences. To avoid these distractions, it is often better to evaluate the consequences of an event on separate dimensions that are combined only when general agreement on the combination rules can be established. These dimensions may differ by domain. By default, TRIAD provides for the assessment of four types of threats: threats to life and health, threats to property, threats to mission (operational) success, and social amplification.

Social amplification refers to the secondary damage caused by people's reactions to an event (Kasperson et al, 1988). This consequence is often underappreciated. For example, the damage caused by a fatal crash of an airliner includes the value of the aircraft, the damage to life, limb, and property in the aircraft and on the ground, and the loss of revenue caused by the loss of the aircraft and the disruption to the schedule. However, the damage caused by a fatal crash of an airliner also includes the psychological trauma endured by survivors and relatives, increases in fears of flying, and damage to the reputation of the airline and the industry. Some of the costs of this damage are borne by the airline or its insurers either directly in payments to individuals or indirectly in lost ticket sales and decreased stock values. Some of these costs are borne by the industry in decreased travel and calls for increased governmental oversight. Some of the costs are borne by the society as a whole. In many cases, the costs associated with social amplification can substantially outweigh all other consequences.

Combining consequences that are assessed on different dimensions presents another problem. Often, the degree of damage will be evaluated on ordinal scales, but the values are treated as if they were interval or ratio scales. This can cause problems. For example, one is tempted to consider a reduction in a consequence rating from "5" to "3" as being greater than a reduction from "3" to "2" although ordinal scales carry no information about the relative sizes of the intervals between the markers. Hence, an intervention that causes a reduction from "5" to "3" may be seen as much more valuable than one that only reduces the rated hazard from "3" to "2". However, because the intervals between categories are not constant, the improvement reflected by a consequence reduction from "3" to "2" may be greater on some absolute scale than the improvement obtained by reducing the rated consequence from "5" to "3" and this latter reduction may be hardly different from a reduction from "5" to "4" (see Figure 1).

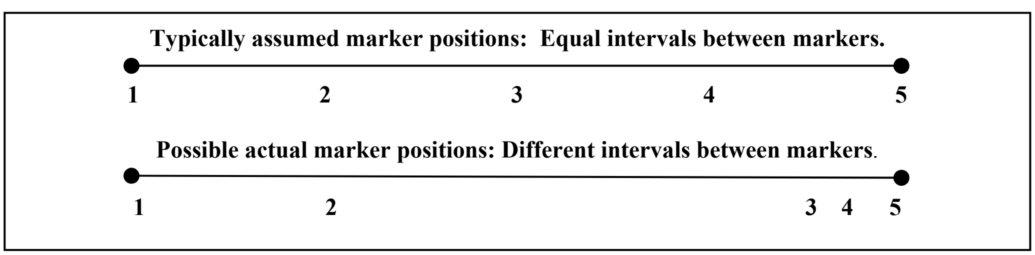

Figure 1. Illustration of possible ordinal values relative to an absolute scale.

This problem is exacerbated when one attempts to combine ordinal scales. Because the same numerals are typically used as markers for relative positions on different scales, users are sorely tempted to treat markers with the same numerical representation as if they were identical and to perform inappropriate arithmetic operations on them. For example, individuals often attempt to multiply the ordinal ratings obtained from two different scales. Consider attempting to combine ratings of threats to life/health and threats to property. If these are made on 5 point scales, the results can be displayed using a matrix like that in Table 1. In general, things get worse from bottom to top, left to right, and along the diagonal from lower left to upper right. However, one cannot easily combine this information into a single summary measure. For example, if the ordinal ratings on each dimension are multiplied, then an outcome rated as Property Damage 5* Life & Health 2 would be considered as risky as an outcome rated Property Damage 2* Life & Health 5 (2*5=5*2=10). But this is not necessarily the case. An incident in which multiple lives are lost but the property damage is $1-$10 million may not be equivalent to one in which there are only minor injuries but the properly damage exceeds $250 million. Neither is it the case that an outcome rated as Property Damage 3* Life & Health 4 (3*4=12) is necessarily worse than one rated as Property Damage 2* Life & Health 5 (2*5=10).

Table 1. *Ordinal Scale Matrix.*

		Property Damage				
		< $ 1 Million	**$1 -$10 Million**	**$10 - $100 Million**	**$100 - $250 Million**	**> $250 Million**
Life & Health		**1**	**2**	**3**	**4**	**5**
Multiple Deaths	**5**					
Single Death	**4**					
Major Injury	**3**					
Minor Injury	**2**					
Minimal/No Effect	**1**					

Extreme Risks

Assessing outcomes with extreme consequences pose a particularly difficult problem (Kunreuther, 2002). In most cases, the traditional calculation of risk as the product of the probability of an event and the potential consequences appears to approximate our sense of what risk is. For example, a business is likely to treat a high likelihood of a small monetary loss as of roughly equivalent risk to a low likelihood of a somewhat larger loss. However, when the probabilities and/or consequences approach their extremes, the risk estimate produced by the traditional calculation departs from what most people feel it should be. In particular, an event that could cause a catastrophe with very low probability is generally seen as much riskier than an event that is highly likely to cause an outcome with very low cost.

This phenomenon is not entirely psychological. Extreme consequences *are* different. For example, an airline can plan for how to respond to most potential outcomes. But one cannot plan for how to respond if the consequence is the collapse of the company. There is a discontinuity in the risk function at the point at which the consequences become unbearable. One cannot treat the collapse of the company, the destruction of an ecosystem, or the death of a society as simply an outcome with very high costs. This does not mean that one cannot assess extreme risks, only that one should not rely on the mechanical application of any simple risk calculation procedure in all situations.

Risk Displays

The value of a risk assessment depends on its ability to inform decisions. Hence, the manner in which the results are displayed is of considerable importance. Risk assessments are often portrayed by a single point on a two dimensional (probability X consequence) display (see left panel, Figure 2). This display neatly summarizes the assessment but it does not provide many important details. From this display, one cannot determine the precision of the assessment. For example, the "+" in Figure 2 may reflect a very precise value or it may indicate a best guess within a 95% confidence interval that extends from 1 to 5. Only one point is displayed (usually the worst case), although a single event may produce several possible outcomes each of which may occur with different likelihoods and cause different consequences. All of the possible types of consequences are combined on a single scale, but the manner in which they are combined is not clear.

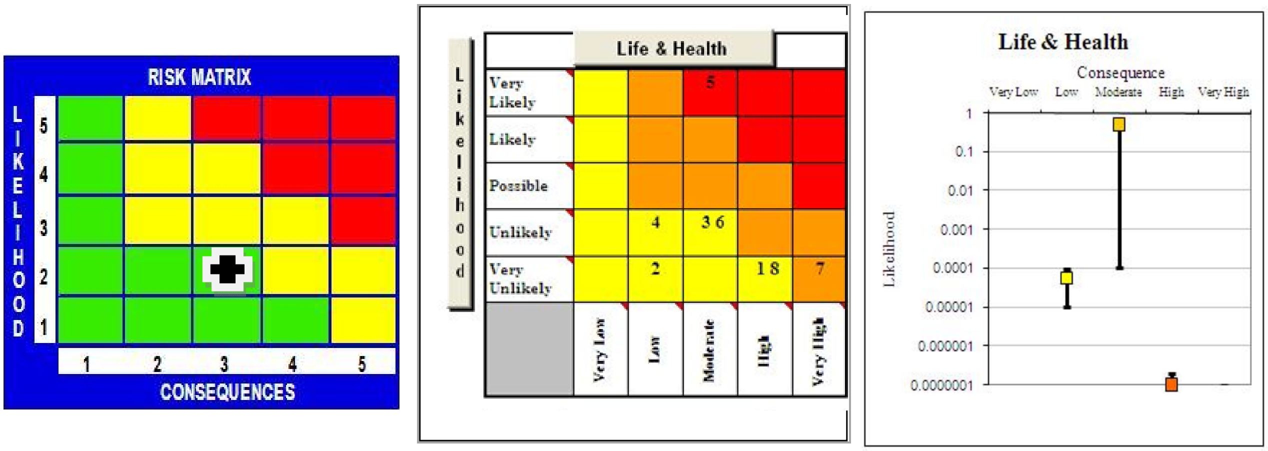

Figure 2. A Common Risk Matrix (left); TRIAD Life & Health likelihood X consequence display (center) and logarithmic risk display (right) showing a possible range of estimates.

In many cases, better decisions may be made if the risks associated with different possible outcomes are displayed, different displays are used for different types of consequences, and confidence intervals around estimates are depicted. TRIAD includes these enhancements (see Figure 2). Different 5 (consequence) X 5 (likelihood) matrices are used to display different consequence dimensions. The evaluators' best estimates of the likelihood and consequence values of each possible outcome (identified by number) are displayed on these graphs (in the center pane of Fig. 2). Auxiliary graphs display the plausible range of likelihood for each outcome (in the right pane of Fig. 2).

Conclusion

In aviation, managers and regulators continually assess risk. However, the heuristics that are commonly used have inherent problems that can render the assessments invalid. Relatively simple steps can be taken to substantially improve the quality of risk assessments even when quantitative data is sparse and traditional probabilistic risk assessment techniques cannot be applied. TRIAD is one tool that can support such comprehensive risk assessment, and can support improved decision making.

Acknowledgements

We thank Nathan Dieckmann, Christina Sheppler, and Ray Vukcevich who assisted with the development of TRIAD.

References

Bier, V. M. & Cox, L. A. (2007). Probabilistic risk analysis for engineered systems. In Edwards, W., Miles, R.F., von Winterfeldt, D. (Eds). *Advances in decision analysis: From foundations to applications*. (pp. 279-301). New York, NY, US: Cambridge University Press.

Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S. (1988). The social amplification of risk: A conceptual framework. *Risk Analysis*. 8, 177-187.

Kunreuther, H. (2002). Risk analysis and risk management in an uncertain world. *Risk Analysis*. 22, 655-664.