ABSTRACT
Like several other transport safety agencies, the Finnish Transport Safety Agency, Trafi, aims to become risk-guided in its decision making and allocation of resources. A two-year project has been set up to support this development.

An important element in the project are the so-called Risk Pictures, where safety data based on safety events is presented on a risk scale. This paper discusses the value of introducing Risk Pictures based on risk metrics in addition to the usual Safety Indicators, as means for Safety/Risk measurement and better guidance for decision making.

The paper takes a conceptual look at Safety Indicators and reviews the current safety indicators of Trafi. While the chosen indicators are well in line with the set criteria, they are bound by the general limitations of Indicators: they are selective and simply count events without assessing event risk – in terms of potential escalation into an accident. It is argued that a risk metric perspective based on systematic event risk assessment is a better base for risk-guided safety decision making than relying only on safety performance indicators. A suitable methodology is outlined and the benefits are illustrated through an example.

Keywords
Safety indicator, risk based oversight, risk metrics, event risk classification, ARMS-methodology, risk picture

TIIVISTELMÄ
Kuten useat muutkin liikenteen turvallisuusvirastot, Trafi pyrkii kohti riskiohjautuvaa päätöksentekoa ja resurssien hallintaa. Tätä varten on käynnistetty kaksivuotinen projektii.


A risk-guided transport safety agency tries to allocate its resources wisely, giving priority to areas and issues which introduce most risk in the transport system. This is a way to maximize safety effectiveness with the available resource. The most common concept today is “Risk Based Oversight” [4, 9] but oversight is only one of the agency’s functions. Functions such as granting operator certificates should not have to be left out of the risk-guided approach.

Especially in aviation, the recent trend has been to move towards a Performance-Based Regulation as opposed to the traditional Compliance Based Regulation [11]. The operator has its own Safety Management System in place and is supposed to find ways to attain the required performance standards without detailed requirements on how to achieve that. One of the consequences is that there may be variable systems in place at operators and the final judgment on whether a particular system produces the required Performance is left to the national transport safety agency.

This increases the desirability and the need to be able to measure the Safety Performance of the Operators and the whole transport system. In modes of transportation where tangible indications of safety – especially accidents – are very rare, this challenge translates into the need to be able to measure Operational Risk levels.

The rational is, that once risk levels are known – or at least estimated – the resource of the agency can be allocated accordingly. An example is the way Transport Canada adjusts its inspection intervals based on perceived air operator risk level [4].

The most natural source for assessing operational risk is the data received from the operations, e.g. in the form of Safety and Operational Reports and accident/incident statistics. The core of the challenge is: how to transform the data set, which is often incomplete and biased, into a meaningful risk picture.

Recent research has taken the standpoint that the classic way to collect and classify safety data (e.g. in aviation) has a low value of predictability to future safety performance [6, page 80]. The modern vision stresses the importance to understand how the operation functions in reality: work-as-done (vs. work-as-imagined) and to try to introduce positive and leading safety measures [10, pages 21-38; and 5, pages 169-187]. Today’s safety data is negative and is supposed to find ways to attain the required Performance is left to the national transport safety agency.

This increases the desirability and the need to be able to measure the Safety Performance of the Operators and the whole transport system. In modes of transportation where tangible indications of safety – especially accidents – are very rare, this challenge translates into the need to be able to measure Operational Risk levels.

The rational is, that once risk levels are known – or at least estimated – the resource of the agency can be allocated accordingly. An example is the way Transport Canada adjusts its inspection intervals based on perceived air operator risk level [4].

The most natural source for assessing operational risk is the data received from the operations, e.g. in the form of Safety and Operational Reports and accident/incident statistics. The core of the challenge is: how to transform the data set, which is often incomplete and biased, into a meaningful risk picture.

Recent research has taken the standpoint that the classic way to collect and classify safety data (e.g. in aviation) has a low value of predictability to future safety performance [6, page 80]. The modern vision stresses the importance to understand how the operation functions in reality: work-as-done (vs. work-as-imagined) and to try to introduce positive and leading safety measures [10, pages 21-38; and 5, pages 169-187]. Today’s safety data is negative and lagging. Therefore, for a Transport Safety Agency willing to subscribe to a modern view of safety and excelling in operational risk measurement, the challenge increases further.

Dekker introduces a mechanism for accidents, where the fatal escalation towards the accident is not the result of a single (or a few) failures, but a slow drift towards failure, where the limits of “normal” and “acceptable” drift over time as a result of cost and other pressures [5]. It is ironical, that one reason for the safety agencies’ growing interest in a risk-guided modus operandi may well be their own pressure to fulfill their increasingly difficult role with resources which seem to be rather diminishing than increasing.

The Finnish Transport Safety Agency

The Finnish Transport Safety Agency, Trafi, was created in 2010, grouping together functions from formerly separate safety agencies for different modes of transportation. Currently, Trafi is the transport safety agency covering Aviation, Maritime, Road and Railway traffic in Finland.

There is an organizational split by mode of transport and by function: Regulation and development; Licenses and approvals; Oversight; Data Resources.

There is also a dedicated Transport Analysis function. Its main task is to analyze information related to transport safety. The results of the analyses are an important input for guiding the whole organization.

In 2012 and 2013 Trafi gradually introduced Safety Indicators. Some aspects of the three-level indicator system are still under development.

There was also a growing desire in Trafi to guide all its activities by risk, i.e. to introduce a risk-guided modus operandi. A two-year project was launched in April 2013 with the aim to define concrete ways in which each of the functions can become risk-guided. The main parts of the project were defined as:

- Review the current safety data sources, data quality/quantity and analysis methods
- Introduce the necessary methods to create risk-based safety metrics (or Risk Pictures).
- Train Trafi staff in the new methods and the underlying safety science
- Identify recommendable ways to improve safety data quality and quantity
- Test new data collection and analysis methods
- Try to introduce new ways to obtain positive safety information, i.e. why things go right
- Review the Safety Measures that Trafi can apply and try to estimate their risk reduction capability for various identified Safety Issues.

The author was tasked to support the project as an external specialist.

This paper discusses the value of introducing Risk Pictures based on risk metrics in addition to the usual Safety Indicators, as means for Safety/Risk measurement and risk-guided decision making.

SAFETY INDICATORS

Since several decades, Safety Indicators have been a popular way to try to reflect the safety performance of an organization or an activity [14]. Øien gives the following definition for an Indicator: “An indicator is a measurable/operational variable that can be used to...”
describe the condition of a broader phenomenon or aspect of reality” [14].

According to Wreathall, indicators are “proxy measures for items identified as important in the underlying model(s) of safety. As such they are uncertain and often only distantly connected to idealized measures that rarely are available in practice” [13]. Wreathall and Jones list the following set of desired characteristics for the selection of indicators [13]:

- Objective. They are based on observable and non-manipulatable sources.
- Quantitative: They are measurable and can identify when changes in performance occur.
- Available: They can be obtained from existing data.
- Simple to understand/represent worthy goals/possess face validity.
- Related to/compatible with other programs.

**Safety Indicators at Trafi**

In the case of the Finnish Transport Safety Agency, the chosen Safety Indicators were designed to reflect the safety level of the Finnish transport system. According to Trafi, the criteria for defining their safety indicators were [7]:

- An indicator must have a known link with transport safety goals and the phenomena under study
- An indicator must be reliable and measurable on a yearly basis (or more often) in a standard way.
- An indicator must be available over long time periods so that improvement in relation to goals can be tracked.

There are three levels of indicators at Trafi. The level 1 indicators are the paramount indicators reflecting performance in terms of accidents, serious incidents and fatalities. Level 2 indicators are organized in terms of “event types which are considered the most frequent causes of accidents and serious incidents” [8]. For aviation, the level 2 indicators are:

- Runway Excursion, RE
- Runway incursion, RI
- Mid-air Collisions and Near Misses, MAC
- Controlled Flight into Terrain, CFIT
- Loss of Control In Air, LOC-I
- Ground Collision, GCOL

According to Trafi, the level 3 indicators facilitate tracking risk factors which could lead to events reflected on level 2. There are over 50 level 3 aviation indicators. Examples include:

- Unstable approach
- Landing gear and thrust reverser related failures
- High speed aborted takeoffs
- Loss of separation due to aircraft
- Incorrect altimeter setting
- Ground Proximity Warning System (GPWS) alerts
- Fire or smoke on board
- Deficiencies in de-icing or anti-icing

**Short Analysis of the Trafi Safety Indicators**

The selected set of Safety Indicators is well in line with the 5 desired characteristics of Wreathall & Jones: they are objective, measurable, available and make sense to aviation professionals. They are also compatible with international practices as the chosen level 1 and 2 categories are in line with industry definitions.

The split between level 2 and level 3 indicators is impossible to make in a fully consistent manner. In an open system like aviation, any contributing factor to an event can be examined and another layer of contributing factors will be discovered – theoretically endlessly [5, page 138]. Therefore, it is not possible to make a clean cut between event types (as pure end states) and causal factors. For example, among the level 2 indicators, CFIT is as close as one can get to an end state while a Runway Incursion as such may be a non-event or an intermediate state followed by a Runway Collision. Similarly, among the level 3 factors many factors would fit well in the role of a contributing/causal/explanatory factor, like “weather” or “fire”, while others fit better in the role of an end state, raising immediate questions about how the situation could escalate that far, like GPWS alerts or Loss of Separation.

Going to the next level of explanatory factors could lead to the identification of factors at a level guiding potential corrective actions more precisely. For example, Situation Awareness, distractions and monitoring/cross-checking (between the captain and the first officer) could be contributing to several level 2 and 3 indicators, like Runway Incursion and Unstable Approach. However, such level of detail would almost certainly not be supported by the collected data and would lead to a very high number of indicators and increased workload in classification.

The split between leading and lagging indicators is similarly somewhat subjective. Timmannsvik [15] defines the two types of indicators as:

- Outcome-based indicators, i.e. Lagging indicators
- Activity-based indicators i.e. Leading indicators

The latter refer to actions taken in a feed-forward manner in an attempt to improve the safety of the system/operation. Examples include [12]:

- Superior provides positive feedback on safety-conscious behavior of the personnel
- Procedures are updated regularly

From this perspective, all the mentioned Safety Indicators at Trafi are Outcome-based Indicators and therefore lagging. Interestingly, Trafi has recently developed a tool for assessing the financial stability of an air operator. The tool covers 25 indicators which are assessed on a color-
coded scale green-yellow-red. These indicators form an entity and the result is analyzed at the level of the tool, rather than individually for each indicator. Therefore, these indicators cannot be considered independent numerical indicators. However, the nature of this indicator set is clearly proactive, reflecting the potential of the organization to perform safely. In the categorization of Reiman & Pietikäinen into Drive, Monitor and Outcome indicators [12], these would clearly fall into the Monitor indicator category.

The author has spent several years supporting airlines setting up and improving their Safety Management Systems (SMS). In that context, he used to categorize indicators in three types:

- Indicators related to safety actions (i.e. identical to Drive Indicators)
- Indicators related to improving the SMS itself (e.g. improving safety data collection through new data sources/increased data quantity/increased data quality)
- Safety Performance indicators (i.e. identical to Outcome indicators)

If Trafi wanted to expand its Indicator system, how could it be done? It does not have easy access to Monitor indicator data, but it could set up indicators related to its own safety actions and related to improving its safety data collection capability, i.e. its visibility to safety in the Finnish transport system. Such indicators could reflect set targets on what Trafi wants to achieve on these two fronts.

Limitations of Safety Indicators

As Grey and Wiedermann state: “the basic inherent difficulty with indicators is that they are selective. They each represent one measure of one aspect of any situation” [14].

It is impossible to be convinced of a good safety level of a complex system by monitoring a relatively small set of parameters. The author argues that if the indicators look bad, then it is quite probable that there are indeed safety issues. However, even if the indicators look good, one cannot trust that the system is safe.

It can also be argued that indicators could work very reliably in a system where the main safety driver is unreliable technology. Frequent component failures would be captured with the indicators and safety measures launched accordingly. Today’s aviation system is characterized by complexity, emergent and transient phenomena, interactions within the socio-technical system and extremely reliable technology. Therefore, outcome indicators might not be an ideal tool for managing aviation safety in today’s context.

From the perspective of this paper, the most important pragmatic limitation of outcome safety indicators is that they simply count how many times certain kinds of events occur - remaining completely insensitive to the content of the event. A loss of separation between two planes without any real risk of collision counts as one. A loss of separation with a less-than-a second margin to a real collision also counts as one. A loss of separation between two small planes each seating one person counts as one. A loss of separation between two large jets totaling 600 souls on board also counts a one.

RISK-BASED SAFETY METRICS

Advantages of Risk-Based Safety Metrics

One way to overcome the limitations of Safety Indicators is to develop Risk-Based Safety Metrics. As (outcome) safety indicators are based on safety event data, the scope here too is limited to event data.

Safety is not about decreasing the number of events, irrespective of the nature of the events. Current safety thinking suggests that over-focusing on fixing all small problems affecting system performance may even be counterproductive in relation to avoiding major catastrophes [2]. Therefore, from a Safety Management perspective, it makes more sense to interpret a set of safety event data through a risk metric than simply as event counts. The main advantage of creating a risk-based safety metric is that it delivers results in this desired format.

The second advantage of risk assessing all events and creating results based on a risk-based safety metric is related to another drawback of the Safety Indicator approach: when Safety Indicators are used, only a subset of all events contribute to the Indicators, and therefore to the safety picture delivered by this approach. The other events are “lost”. This drawback should be of concern latest when one considers that the Safety Indicators have been prescribed before-the-fact and may therefore become a self-fulfilling-prophecy. In the method described below, all events contribute to the safety picture. Among other things, this may lead to discovering unexpected patterns.

Finally, when the event risk assessment is carried out correctly, event risk becomes another property of each single event and depends only on the individual event. Therefore, event risks can be summed together. This way, total risk values can be obtained for any given entity, for example total risk of all events in February, risk of all events on a particular aircraft type, and so on. Such risk sums can provide very useful risk pictures with different perspectives to the operation. Any descriptors associated with the events can be used to construct different kinds of risk pictures, e.g. month, year, time, location, aircraft type, phase of flight, operator, type of operation, etc. An example of a risk picture constructed by using event risk sums by month is presented in figure 1.
Running a system of indicators is relatively easy. For each new event, one must check whether it contributes to any safety indicators. If it does, the count of those indicators increase by one. If one wants to go beyond indicators and start looking at the risk picture drawn collectively by all the reported events, each individual event needs to be risk-assessed. To do this properly, one must consider the possibility that the event would have escalated into an accident, and try to estimate both the severity of such an accident; and the likelihood of such an escalation. The method for event risk assessment will be covered in the following chapter.

**The Method**

This chapter outlines a method for creating risk pictures based on summing together event risks. The method is essentially the ARMS methodology, created by the Airline Risk Management Solutions (ARMS) working group 2007-2010, which the author chaired. A full description of the ARMS methodology and the working group is available at Skybrary [1]. Review of the ARMS methodology in today’s safety management context will be the topic of a separate dedicated paper by the author.

The experience of the ARMS working group was that risk assessment was typically done using the likelihood \(x\) severity formula without stopping to think of the fundamental questions on what is actually risk assessed, what does the likelihood refer to and what does the severity refer to. One of the key paradoxes was that risk is supposed to contain uncertainty about the outcome (otherwise it would already be a loss, not a risk), and past events do not contain uncertainty as they already took place and the outcomes are known. So how can one risk assess a past event and how should the result be interpreted? From the practical perspective this question is of paramount importance, as airlines typically have access to a lot of safety data, and they would like to use it to obtain risk information.

Perhaps the most valuable contribution of the ARMS work was to answer this and other similar fundamental conceptual questions related to risk assessment in an operational context. Laying the conceptual foundation is a vital step before one should even try to create any practical applications such as risk matrices.

One of the most important conclusions from the working group was that there are two different risk assessment tasks which need to be kept apart: risk assessing a single event and risk assessing Safety Issues (typically reflected through several events). The traditional way has been to try to apply the likelihood \(x\) severity formula for single events and often the former term has been interpreted as likelihood of recurrence or frequency of similar events. This approach introduces a conceptual problem, because in this case likelihood is not a property of a single event. Logically, the risk of a single event should depend only on that single event.

According to ARMS, single event risk is the risk that instead of the actual outcome, the event would have escalated into an accident. The risk then reflects both the probability of the escalation and the severity of the accident outcome. This definition addresses the above paradox about the lack of uncertainty when dealing with past events. The assessment is guided by the following logic:

1. The analyst needs to create a scenario for the escalation. Safety work is done in relation to accidents and consequently risk assessments also need to be done in relation to accidents. Therefore, the scenario needs to include an escalation into an accident outcome (and not only into some intermediate consequence, like flight cancellation). The accident should be the most likely accident type starting from the actual event, in the actual circumstances of the day (e.g. weather etc.). For example, a loss of separation event could have escalated into a midair collision.

2. Based on the chosen scenario, the severity of the accident outcome can be derived from the resulting accident type.

3. The probability dimension now represents the probability of the escalation. If one wants to reason in terms of barriers, this translates into the combined estimated effectiveness of the remaining barriers.

Defined this way, the event risk only depends on the event itself. The fact that the “frequency of similar events” is not included in the formula means that such event risks are separate independent values and can be summed together to obtain the cumulative risk for a group of events.

There is an abundance of detail concerning this type of event risk assessment. Such discussion goes outside the
scope of the current paper but will be covered in a dedicated paper.

It is now possible to create charts which look very similar to charts reflecting the values of Safety Indicators – except that now the results are presented on a risk metric. One interesting experiment is to take the same set of events which contributes to some chosen indicators and create the corresponding risk picture. This gives the opportunity to compare the results: does the result look different on the risk metric?

Such an example is developed in the following chapter.

Example – Same Data Set Seen Through Safety Indicators vs. Risk Metric

The example data set consists of 381 safety reports made by pilots guiding in ships on the Finnish coast. Figure 2 presents a typical indicator view to the results. The events have been categorized to get an idea of the key factors having contributed to the events. For the purposes of this example, the focus is on the top two categories: “Ship maneuvering skills” and “Technical failure”. In this figure, the results have been presented as a count of events in each category. The category “Technical failure” seems to be the most important issue to focus on for safety improvement.

Figure 2. Ship pilot data presented as Safety Indicators reflecting key issues in events during ship pilotage.

Figure 3 presents the same data set using a relative risk metric. The underlying event risk classification has been carried out using the ARMS methodology but with the severity classes of the International Maritime Organization (IMO):

- Very serious casualty to ships.
- Serious casualty to ships.
- Less serious casualty to ships.
- Marine incident.

The risk scale is relative, so the absolute values have no meaning. However, it can be observed that in terms of risk, the Technical failures have lost roughly half of their amplitude compared to Ship maneuvering. This means that there are a lot of events involving technical failures but such events do not have a high (average or cumulative) risk potential (in terms of escalation into accident and severity of such potential accident). Compared to that, the Ship maneuvering events produce a lot more operational risk. Additionally, in this category there are “red events” i.e. events which have got a very high risk class individually. There are no such events in the Technical failure category.

Figure 3. Same data set as in figure 2 but presented on a scale of relative risk, i.e. using a risk metric.

Looking at one or the other way to present the results would guide decision making in a different way. This exactly is the point of the present paper: looking solely at event counts may be a misleading way to allocate safety improvement resource. Looking at the results through a risk metric should provide more appropriate guidance for safety decision making.

DISCUSSION

While this paper supports the presented approach using event risk and safety metrics, there are several limitations that should be kept in mind.

The event risk assessment method uses a relatively simple model to describe the potential escalation that could have taken the event to an accident outcome. Barriers and barrier thinking are one aspect of the model. Such an approach falls short of the latest thinking in terms of safety and accident causation [5; 10]. Because of this, it can be assumed that this approach would address mainly the events within the common-cause variation but not the events related to special-cause variation [3].

When considering the effectiveness of the remaining barriers, there is no bowtie-type method proposed for deriving the total effectiveness based on the individual barriers. This can be seen either a limitation or a strength based on how barrier analysis is understood. If one believes that the total barrier effectiveness can be derived from the individual barriers through some relatively simple operations, then lacking such guidance is a limitation. If one believes that the barrier system contains numerous non-linear interactions and may be complex, then it could be
argued that expert judgment of the total effectiveness is at least as good as a method based on reductionism.

Carrying out an event risk assessment on all incoming events requires some extra work. On the other hand, taking the analysis work seriously means that each event needs to be read through (and classified) in any case. Experience has shown that once the event has been read, the event risk assessment is very quick, typically around 30 seconds.

Finally, the data itself creates important limitations. The results are valid only inside the given data set and one has to be careful when trying to extrapolate the results. Even more importantly, every data source tends to be biased one way or another, and being blind to such biases could lead to gross misinterpretations of the results. For example, safety reporting will almost certainly filter out cases where rules or procedures are violated. Flight data would reveal some of the violations very reliably but be completely blind to communication issues.

In defense of the risk metric approach, one should point out that most of the mentioned limitations are not specific to the risk metric approach but rather general limitations, also applicable to the Safety Indicator approach. The author considers the risk pictures to bring a significant added value to the information provided by Safety Performance Indicators. The author also considers the ARMS approach for carrying out event risk assessments more robust and reliable than older methods based on the severity-frequency dimensions.

The purpose of this paper has been to introduce the risk metric approach and to illustrate its potential in supporting risk-guided safety decision making. The 2-year project at Trafi will aim at creating risk metrics which are common to the four modes of transport and use other types of safety information (e.g. audit results) to create integrated results where biases can be better controlled. There is also an aim to bring the latest safety thinking to the everyday methods and tools at the Transport Safety Agency.

CONCLUSIONS

This paper has discussed two approaches to safety performance monitoring in view to guiding safety decision making. The first is based on Safety Performance Indicators. The second is based on giving each individual event a value of event risk, and using such data to construct a risk picture where the results are presented using a (relative) risk metric.

The point was made that indicators reflect event counts and are completely insensitive to the real accident potential of individual events. Additionally, information through Safety Performance Indicators is produced only by the subset of events which fall under the prescribed indicator categories.

Using event risk assessment and presenting the results on a risk metric takes advantage of the full data set and measures risk instead of event count. It is argued that this is a better base for risk-guided safety decision making than relying only on safety performance indicators.

The ARMS methodology presents the conceptual basis for carrying out event risk assessment and the methodology can be customized for specific needs.
REFERENCES


