
How does aviation industry measure safety performance? Current practice and limitations

Steffen Kaspers*

Aviation Academy,
Amsterdam University of Applied Sciences,
Weesperzijde 190, 1097 DZ Amsterdam, the Netherlands
and
Faculty of Aerospace Engineering,
Delft University of Technology,
Kluyverweg 1, 2629 HS Delft, the Netherlands
Email: s.e.kaspers@hva.nl
*Corresponding author

Nektarios Karanikas

Aviation Academy,
Amsterdam University of Applied Sciences,
Weesperzijde 190, 1097 DZ Amsterdam, the Netherlands
Email: n.karanikas@hva.nl

Alfred Roelen

Aviation Academy,
Amsterdam University of Applied Sciences,
Weesperzijde 190, 1097 DZ Amsterdam, the Netherlands
and
Netherlands Aerospace Centre (NLR),
Anthony Fokkerweg 2, 1059 CM Amsterdam, the Netherlands
Email: alfred.roelen@nlr.nl

Selma Piric and Robert J. de Boer

Aviation Academy,
Amsterdam University of Applied Sciences,
Weesperzijde 190, 1097 DZ Amsterdam, the Netherlands
Email: s.piric@hva.nl
Email: rj.de.boer@hva.nl

Abstract: This paper reviewed state-of-the-art literature about existing aviation safety metrics. We identified that the long-established view on safety as absence of losses has limited the measurement of safety performance to indicators of adverse events. However, taking into account the sparsity of incidents and accidents compared to the amount of aviation operations and the recent shift from compliance to performance based approach to safety management, the exclusive use of outcomes metrics does not suffice to further

improve safety. Although the need to use activity indicators is recognised, those have not yet become part of safety performance assessment. This is partly attributed to the lack of empirical evidence about the relation between safety proxies and outcomes and the diversity of safety models used (i.e., root-cause, epidemiological or systemic models). This has resulted to the development of many safety process metrics, which have not been thoroughly tested against the quality criteria.

Keywords: safety management; safety performance; safety indicators; safety; aviation.

Reference to this paper should be made as follows: Kaspers, S., Karanikas, N., Roelen, A., Piric, S. and de Boer, R.J. (2019) 'How does aviation industry measure safety performance? Current practice and limitations', *Int. J. Aviation Management*, Vol. 4, No. 3, pp.224–245.

Biographical notes: Steffen Kaspers started his career in the military as an officer fighter control in the Royal Netherlands Airforce. During his training, he became interested in topics such as 'situational awareness' and safety. After the military he followed a Master Human Factors and System Safety in Lund, Sweden. With that knowledge he worked as a consultant applying theory in practice. Currently, he combines working as a researcher and teacher at the Amsterdam University of Applied Sciences, pursuing a PhD on the topic of Measuring Safety in Aviation.

Nektarios Karanikas is an Associate Professor of Safety and Human Factors at the Aviation Academy. He studied human factors and safety assessment in aeronautics at Cranfield University and was earned his Doctorate in Safety and Quality Management from Middlesex University. He graduated from the Hellenic Air Force Academy as Aeronautical Engineer, worked for 18 years as officer and resigned in 2014 with the rank of Lt. Colonel. He served in various positions related to maintenance and quality management and accident investigation and he was lecturer and instructor for safety and human factors courses.

Alfred Roelen is an expert in system safety. He obtained his MSc in Aeronautical Engineering from the Delft University of Technology in 1992. From 1992 to 1994, he joined the Technological Designer program at the Delft University of Technology, Faculty of Aerospace Engineering, obtaining a Master of Technological Design (MTD) in 1994. In 1994, he joined the National Aerospace Centre NLR, Department of Flight Testing and Safety. He is currently a Senior Scientist at the NLR's Air Transport Safety Institute and is a researcher-Lecturer at the Aviation Academy for one day a week.

Selma Piric is a Lecturer and researcher at the Amsterdam University of Applied Sciences. As an aviation consultant, she has worked for various air navigations service providers throughout the world. Her key areas of expertise include ATM operations, validation, simulation, safety management systems, safety culture, aviation safety, logistics and optimisation, airport planning and SESAR. She holds an MSc in Human Factors in Aviation Safety from the Cranfield University and a BSc in Organisational Psychology.

Robert J. de Boer is a Professor of Aviation Engineering at the Aviation Academy. Robert's key research interest is Human Performance in socio-technical systems. He was trained as an Aerospace Engineer at the Delft University of Technology, majoring in man-machine systems and graduating cum laude in 1988. After gaining experience in line management and

consulting, he joined Fokker Technologies in 1999, where he developed the program management methodology for Fokker before being appointed Director of Engineering in 2002. These experiences inspired his current scientific interest in team collaboration, cumulating in a PhD at the Delft University of Technology.

This paper is a revised and expanded version of a paper entitled 'How does aviation industry measure safety performance? – Current situation, limitations and next steps' presented at 20th ATRS World Conference in Rhodes, Greece, 23–26 June 2016.

1 Introduction

The improvement of aviation safety has been a focal point for companies and authorities. Several accidents led to specific improvements for aircraft, for example, the cockpit voice recorder and flight data recorder were introduced after a crash in the 60's, crew resource management was introduced after Tenerife and also the Airborne Collision Avoidance System and Ground Proximity Warning Systems were introduced after specific accidents. Three key areas could be identified in the evolution of safety: the technical era, the human factors era and the organisational era (ICAO, 2013b). Where the rate was around 50 accidents per million departures in 1960 and around 5 per million departures in the eighties, currently it has improved to around three accidents per million departures despite the growing traffic (Boeing, 2016).

Although statistics show low rates of accidents (e.g., ICAO, 2016), safety remains challenged in daily operations as indicated by the safety data collected through various initiatives (e.g., voluntary reports, flight data monitoring, audits). To further improve safety, new international and regional guidance and regulations for safety management have been set (e.g., ICAO, 2013a; FAA, 2013; EASA, 2014; EC, 2014). Those differ significantly from conventional quality assurance, which has been emphasising on the presence and operation of a process (i.e., compliance-based assessment) and add the requirement for monitoring safety performance through relevant metrics (i.e., performance-based assessment). This new approach is expected to render safety management more proactive than it is currently since proper monitoring of the safety management activities and outcomes of those will allow identifying and managing flaws before accidents occur.

Proactive safety relies heavily on the use of relevant data from day-to-day activities and operations; the concept is that processing of such data will allow timely identification and control of new/changed hazards and combinations of those, the deviations from standards included. However, although large companies might be in place to collect adequate amounts of safety-related data and establish proactive safety metrics, small and medium enterprises (SMEs) lack high volumes of such data due to the limited scope of their operations. Furthermore, even at the case of large companies, more reactive indicators are in use than proactive ones (e.g., Woods et al., 2015; Lofquist, 2010) and considerable amount of resources are required to process high volumes of operational data (e.g., Kruijssen, 2013).

Taking into account the situation described above, the Aviation Academy of the Amsterdam University of Applied Sciences initiated a 4-years' research project entitled

Measuring Safety in Aviation – Developing Metrics for Safety Management Systems. The aim of the project is to identify ways to measure safety levels without the benefit of large amounts of safety data (Aviation Academy, 2014). The researchers will develop and validate new safety metrics based on various approaches to safety and translate this knowledge into a web-based dashboard for the industry. The project was launched in September 2015, is co-funded by the Nationaal Regieorgaan Praktijkgericht Onderzoek SIA (SIA, 2015) and is executed by a team of researchers from the Aviation Academy in collaboration with a consortium of representatives from the industry, academia, research institutions and authorities.

As part of the aforementioned project, this paper reviews current suggestions and practice in safety metrics. State-of-art academic literature, (aviation) industry practice and documentation published by regulatory and international aviation bodies were considered in this review. The criterion used for selecting academic references was the date of publication (i.e., about up to 10 years old) and relevance to the topic (keywords: safety metrics, safety indicators, safety performance); the main online repositories consulted were ScienceDirect and Google Scholar. The most current versions of relevant aviation standards and guidance were reviewed along with information presented by companies during public events (e.g., conferences, symposia).

Although this review is about safety metrics, the latter cannot be viewed outside their context. Hence, this paper starts with presenting various views on safety and the challenges when measuring safety. Next, we present various positions and approaches regarding safety metrics, the role of safety management systems (SMS) in safety monitoring, classifications of safety performance indicators (SPI) and the quality criteria of metrics in general. The paper continues with a discussion on the literature and industry references reviewed and presents the respective conclusions.

2 Review of literature and industry references

2.1 Views on safety

2.1.1 Long-established views on safety and relevant limitations

The International Standardization Organization (ISO) defines safety as “...freedom from unacceptable risk...”, risk as “...a combination of the probability of occurrence of harm and the severity of the harm...” and harm as “...physical injury or damage to the health of people either directly or indirectly as a result of damage to property or the environment” (ISO, 1999). International Civil Aviation Organization (ICAO) defines safety as: “...the state in which the possibility of harm to persons or of property damage is reduced to and maintained at or below, an acceptable level through a continuing process of hazard identification and safety risk management” [ICAO, (2013b), pp.2–1]. Both definitions include the term risk, which is defined as a combination of probability and severity of harm and they refer to acceptable levels of risk, thus suggesting the existence of a threshold between safe and unsafe states.

The aforementioned views on risk are linked to a deterministic approach to safety, where probabilities can be determined either quantitatively based on frequencies of past events, or qualitatively through expert judgment, the latter including various limitations due to the influence of heuristics and biases (Duijm, 2015; Hubbard and Evans, 2010).

Likewise, the severity of harm is estimated through credible accident scenarios (ICAO, 2013b), which are based on extrapolation of previous experience and the assumption that the set of accident scenarios is finite and available. This can apply to general categories of events (e.g., controlled flights into terrain, runway excursions) but it is not always feasible when considering combinations of various factors that can contribute in those high-level events (Roelen and Klompstra, 2012; Leveson 2011).

Table 1 Various definitions of safety

“...freedom from unacceptable risk...”	ISO (1999)
“...the state in which the possibility of harm to persons or of property damage is reduced to and maintained at or below, an acceptable level through a continuing process of hazard identification and safety risk management”	ICAO (2013b)
“...a dynamic non-event...”	Weick and Sutcliffe (2001)
emergent behaviour or property of complex systems	Dekker et al. (2011), Cilliers (1998), Dekker (2011) and Leveson (2011)
a product of complex interactions that can be explained after an event, but their effects on normal operations were not fully understood before the event	Snowden and Boone (2007)
the ability of a system to achieve its objectives under varying conditions.	Hollnagel (2014)

The definitions of harm in relation to safety exclude acts of terrorism, suicide or sabotage such as the recent losses of Germanwings and MetroJet aircraft (Flight Global, 2016). The levels of other types of operational risks that are calculated via a risk assessment process must be compared against what is acceptable, so to identify whether mitigation is required.

However, the level of acceptable operational risk has not been universally established; ICAO (2013b) prompts States and organisations to define their own risk tolerances and thresholds, rendering thus cumbersome to make respective comparisons across the aviation industry. Furthermore, the acceptability of risk depends on the system considered; a single fatality can be perceived as a big loss at company or individual level, but might not be seen as such at the level of State or industry sector (Papadimitriou et al., 2013; Pasman and Rogers, 2014; Sinelnikov et al., 2015). For example, Ale (2005) suggested a maximum acceptable individual fatality risk of 1×10^{-6} per year in the Netherlands, but also identified a strong sensitivity of the public to multiple fatalities resulting from a single event.

Furthermore, international, national and professional group norms and cultures may influence acceptable risks (ICAO 2013b), while the perception of safety might differ from the officially accepted risk levels. In practice, the sense of safety is often eradicated in the wake of adverse events which mandate actions to prevent reoccurrence, regardless the maintenance of acceptable risk levels (Dekker, 2014). Also, the occurrence of a harmful event may signal that the a-priori probabilities were estimated too optimistically (Hopkins, 2012), or that the organisation might over time have overweighed productivity and efficiency in the expense of safety, as the lack of the latter can be evident through rates of safety occurrences attributed to human error (Karanikas, 2015a).

2.1.2 Alternative views on safety

Weick and Sutcliffe (2001, p.30) defined safety as "...a dynamic non-event...". The authors stressed out that we recognise safety by the absence of harm (i.e., something bad not happening) in a constantly changing context, so we actually define safety through non-safety. Various authors (e.g., Dekker et al., 2011; Cilliers, 1998; Dekker, 2011 cited in Salmon et al., 2012; Leveson, 2011) viewed safety as emergent behaviour or property of complex systems. Under this approach, safety is a product of complex interactions that can be explained after an event, but their effects on normal operations were not fully understood before the event (Snowden and Boone, 2007). Thus, as Lofquist (2010) argued, there is a need to consider interactivity within socio-technical systems when measuring safety.

Hollnagel (2014) introduced the concept of Safety-II, where safety is defined as the ability of a system to achieve its objectives under varying conditions. The aforementioned author claimed that both desired and unwanted outcomes derive from the same human and system behaviours, called performance adjustments and that the variability of outcomes is a result of complex interactions of system elements rather than failures of single components. Based on a similar thinking, Grote (2012) concluded that contingencies need to be part of safety management activities so the system will be able to respond successfully to variances and disturbances; Perrin (2014) proposed the use of success-based metrics in safety assessment.

2.2 Safety performance metrics

Safety management regards the activities and processes for achieving safety goals in a systematic manner and can be interpreted as a set of organisational controls for safety (Wahlström and Rollenhagen, 2014). In the safety assurance pillar of SMS, the monitoring of safety indicators and the assessment of safety performance are prescribed; appropriate targets need to be set for SPI in the frame of an SMS (UK CAA, 2011; Holt, 2014). According to ICAO (2013a, pp.1–2) safety performance is "A State or a service provider's safety achievement as defined by its safety performance targets and SPI", "[Safety performance indicator:] A data-based parameter used for monitoring and assessing safety performance" and "[Safety performance target:] the planned or intended objective for safety performance indicator(s) over a given period".

ICAO (2013b) describes indicators at two levels, the State level, which monitors its safety indicators and the individual service provider that monitors SPI as part of its SMS. Within the SMS, another distinction is made: high consequence indicators refer to accidents and serious incidents (e.g., air operator's monthly serious incident rate); low consequence indicators are based on activities and incidents (e.g., voluntary reports received annually). In aviation, accidents are defined as events "...associated with the operation of an aircraft [...] in which a person is fatally or seriously injured [...], the aircraft sustains damage or structural failure [...], or the aircraft is missing or is completely inaccessible" [EC, (2014), p.L122/25]. Also, the European Commission [EC, (2014), p. L122/25] considers as occurrence "...any safety-related event which endangers or which, if not corrected or addressed, could endanger an aircraft, its occupants or any other person...". Each occurrence is classified as (EC, 2010; ICAO, 2010):

- Incident: an occurrence, other than an accident, associated with the operation of an aircraft which affects or could affect the safety of operation.
- Serious incident: an incident involving circumstances indicating that there was a high probability of an accident. Hence, the difference between an accident and a serious incident lies only on the actual result.

In the safety performance assessment tool created by the Safety Management International Collaboration Group (SMICG, 2013), metrics are divided into three tiers, where tier 1 metrics measure the outcomes of the whole civil aviation domain, tier 2 indicators depict safety management performance of operators and tier 3 metrics address the activities of the regulator (SMICG, 2014). SPI should have an alert level (i.e., limit of what is acceptable) and safety indicators support the monitoring of existing risks, developing risks and implementation of mitigation measures (ICAO, 2013b). If implemented in this way, safety management allows a performance-based approach, which is expected to create more flexibility for the users to achieve safety goals in addition to compliance. SPI might have up to three functions within safety management: monitoring the state of a system, deciding when and where to take actions and motivating people to do so (EUROCONTROL, 2009; Hale, 2009); their establishment may also foster motivation towards safety (Hale et al., 2010).

Also, safety management is often linked to safety culture (e.g., Stolzer et al., 2008), the latter lacking a common definition in the literature (Guldenmund, 2007) and being assessed with various instruments which in their majority lack external validity and do not support understanding of the survey findings (Karanikas, 2016). The European Union's Single European Sky Performance Scheme added the assessment of Just Culture within an organisation as a leading indicator (EUROCONTROL, 2009). However, the literature is not aligned regarding the view whether safety culture is a result of safety management, thus a type of outcome indicator, or a reflection and indication of how well safety management is performed (Piric, 2011).

2.3 Classification of safety performance metrics

In professional and scientific literature SPI are often classified as 'lagging' or 'leading'. Grabowski (2007, p.1017) stated: "Leading indicators, one type of accident precursor, are conditions, events or measures that precede an undesirable event and that have some value in predicting the arrival of the event, whether it is an accident, incident, near miss, or undesirable safety state. [...] Lagging indicators, in contrast, are measures of a system that are taken after events, which measure outcomes and occurrences". According to SMICG (2013), lagging indicators are safety outcome metrics since they measure safety events that have already occurred, whereas leading indicators can be used to prioritise safety management activities and determine actions towards safety improvement.

Harms-Ringdahl (2009) proposed the use of the terms activity and outcome indicators in correspondence with leading and lagging indicators. Reiman and Pietikäinen (2012) made a distinction within leading indicators: driving indicators facilitate aspects within the system and they measure safety management activities (e.g., independent safety reviews and audits are carried out regularly and proactively); monitoring indicators measure the results of driving indicators (e.g., the findings from external audits concerning hazards that have not been perceived by personnel/management previously). Hollnagel (2012, p.4) proposed two types of indicators: reactive indicators "...keeping an

eye on what happens and to make the necessary adjustments if it turns out that either the direction or the speed of developments are different from what they should be...” and proactive indicators “...to manage by adjustments based on the prediction that something is going to happen, but before it actually happens...”.

From a process safety perspective, Erikson (2009) suggested that leading indicators correspond to inputs and lagging indicators to outputs, thus all indicators might be characterised as both leading and lagging depending on their place in the process. Øien et al. (2011a) defined both risk and safety indicators as leading indicators: risk indicators are metrics based on and tied with the risk model used for assessing the level of safety and measure variations of risk levels; safety indicators do not need to refer to an underlying risk model, they can stem from different approaches such as resilience based (e.g., Woods, 2006), incident based, or performance based ones, but they should still be measurable.

Table 2 Pairs of indicator types referred in literature

<i>Classifications of safety performance indicators</i>	
Leading	Lagging
Upstream	Downstream
Predictive	Historical
Heading	Trailing
Positive	Negative
Active	Reactive
Predictive	Retrospective
Input	Output
Driving/monitoring	Lagging
Proactive	Reactive
Activity	Outcome

In an attempt for a more elaborate classification than simply leading and lagging, Hinze et al. (2013) suggested the distinction of safety leading indicators to passive and active. Passive leading indicators address the state of safety in long term or a macro scale (e.g., a requirement that each subcontractor submit a site-specific safety program that must be approved prior to the performance of any work by that subcontractor). Active leading indicators represent safety in short term (e.g., percent of jobsite pre-task planning meetings attended by job-site supervisors/managers, number of close calls reported per 200,000 hours of exposure). Hale (2009) addressed the confusion about leading and lagging indicators and attributed this to variances in:

- 1 the ‘degree’ of leading
- 2 compression of the temporal dimension
- 3 the categorisation of causal factors (e.g., unsafe acts, unsafe conditions).

Table 2 shows the various classifications discussed by the authors cited in this section.

2.4 *Safety outcome metrics*

As ICAO (2013a) and the European Commission (EC, 2014) mention, the reporting of (serious) incidents primarily aims at finding ways to improve safety rather than depicting safety performance. This leaves only accidents as indications of safety performance, as reflected in the annual safety reports published by various organisations (e.g., ICAO, 2016; Flight Global, 2016; IATA, 2016; Boeing, 2016). Those refer mainly to accident data segregated for regions, aircraft size, types of operation etc.; apart from raw numbers, ratios of safety events by activity units are calculated (e.g., number of flights and departures, flight hours, passenger miles) to facilitate comparable measurements of safety performance. Various authors (e.g., Bourne et al., 2013; Singh et al., 2012) recognised that association of performance with actual results of processes is widely accepted.

However, the aforementioned practice contradicts with the view that safety performance is monitored based both on outcome and activity data and that incidents, serious incidents and accidents are collectively considered as outcomes (ICAO, 2013b). Hence, the use of the term safety outcome metrics must ideally refer to occurrences of all severity classes. Some recent proposals for safety performance metrics and assessment methods, which are based on all outcomes regardless severity, are presented below:

- Di Gravio et al. (2015) proposed a statistical analysis of safety events based on Monte Carlo simulation and weighing of factors, as means to develop proactive safety indicators.
- Karanikas (2015b) suggested to consider the extent to which an event was controllable and the end-user attempted to positively intervene in the unfolding situation the mere reference to severity of events without prior considering their potential might not be representative of safety performance.
- Bödecker (2013) claimed that safety performance can be measured through consideration of frequencies and risk levels of events identified from occurrence reports and audit findings.
- ARMS Working Group (2010) proposed the assessment of safety performance through a combination of event risk classification (ERC) values, based on a risk matrix and safety issue risk assessment (SIRA) values based on a bow tie diagram.
- The aerospace performance factor (APF) tool used by the EUROCONTROL (2009) maps an overall safety trend based on outcomes and reflecting the relative risk over time.

2.5 *Safety process metrics*

The fact that safety occurrences are sparse compared to the amount of operational activities does not allow to timely monitor safety performance variations and distance of operations from unacceptable risks (Espig, 2013; O'Connor et al., 2011). According to Espig (2013) "...[we] need measures of our performance based on how we deliver a safe service, day-in, day-out, to inform us of our performance variation and our 'distance' from the incident". Therefore, other types of metrics have been suggested as proxies for safety performance (Wreathall, 2009). Those metrics offer indirect indications of safety performance and can be used as early warnings of accidents (Øien et al., 2011). In this

paper, we refer to those as safety process metrics in order to distinguish them from safety outcome metrics (see Section 2.4).

The predictive power or validity of safety process metrics has to be demonstrated through empirical evidence or inferred through credible reasoning (Wreathall, 2009). However, there is limited scientific evidence for the relation between safety outcome metrics and safety process metrics (Reiman and Pietikäinen 2012), although in other industries such as occupational and process safety, evidence was found (e.g., Lardner et al., 2011). Therefore, the validity of process metrics in aviation is mostly dependent on credible reasoning, the latter reflecting the application of specific safety models (Wreathall, 2009).

Three families of safety models can be found in the literature:

- Single (root) cause models, such as the ‘domino’ model, which suggest that a triggering event sets a causal sequence in motion that leads to a harmful event (e.g., Underwood and Waterson, 2013).
- Epidemiological (multiple causes) models, such as the ‘Swiss cheese’ model (Reason, 1990), which differentiate between active failures (i.e., actions and inactions) and latent conditions (i.e., individual, interpersonal, environmental, supervisory and organisational factors present before the accident) that jointly lead to a harmful event. In the aviation domain, the specific model has been included in the safety management manual (ICAO, 2013). The use of defences to counteract for possible failures is common across those types of models, such as the bow-tie (e.g., Boishu, 2014), threat and error management (e.g., Maurino, 2005) and Tripod (e.g., Kjellen, 2000).
- Systemic models such as STAMP (Leveson, 2011), FRAM (Hollnagel, 2010) and Accimap (Rasmussen, 1997) that focus on component interactions rather than single component failures in a dynamic, variable and interactive operational context. Although ICAO (2013b) views SMS as a system, a respective consideration of dependencies and linkages of system components is not visible.

2.6 *Quality of metrics*

Karanikas (2016) discussed the limited practicality, validity and ethicality of some safety metrics proposed in literature or applied in practice. Various authors mentioned quality criteria for indicators, addressing though that it is difficult to develop indicators that fulfil all requirements and that, in practice, it is even challenging to judge to what extent a metric meets each criterion (Hale, 2009; Hinze et al., 2013; Karanikas, 2016; Podgórski 2015; Sinelnikov et al., 2015; Webb 2009; Øien et al., 2011a, 2011b; Rockwell, 1959). The following list refers to the quality criteria the aforementioned authors suggested:

- based on a thorough theoretical framework
- specific in what is measured
- measurable, so to permit statistical calculations
- valid (i.e., meaningful representation of what is measured)
- immune to manipulation

- manageable – practical (i.e., comprehension of metrics by the ones who will use them)
- reliable, so to ensure minimum variability of measurements under similar conditions
- sensitive to changes in conditions
- cost-effective, by considering the required resources.

Saracino et al. (2015) and Tump (2014) proposed that metrics are more useful when their purpose and context are clear, by considering:

- What the indicator targets to measure.
- The context and area that the indicator belongs (e.g., size of the company, type of activities such as air operations, maintenance, ground services, air traffic management).
- What type of hard or/and soft data are required and how the latter will be quantified.
- Control limits for monitoring the calculated values.
- What laws, rules and other requirements the indicator might fulfil.

3 Discussions

Following the review of literature and industry practice, first we noted that the definition of ISO limits safety to the lack of “...physical injury or damage to the health of people...”, either directly, or indirectly incurred through damage to property or the environment. ICAO, on the other hand, in addition to harm on people, includes any type of damage as non-safety. Also, ICAO views safety as a state where acceptable risk levels have been achieved “...through a continuing process of hazard identification and safety risk management...”, thus implying that safety is a state that needs to be maintained through a risk management process such the one introduced in SMS. The relation between risk (i.e., probability of harm) and safety (i.e., level of risk) means that a system may have been in an unsafe state even though no visible harm has been experienced (i.e., accidents) and, reversibly, a system can be considered safe even though harm was experienced, because the overall risk level might still be in the acceptable area. This approach actually matches the state-of-the-art thinking on complex systems, which suggests that continuous control loops and monitoring are required to maintain a system within predefined safety boundaries. However, despite newer views of safety have been articulated (e.g., emergent property of complex systems) and modern safety models have been developed (e.g., STAMP, FRAM), the long-established view of safety as a risk of harm and the epidemiological models are mostly recognised in industry standards.

The current classification of incidents as serious or not does not draw clear lines, whilst non-standardised terms are used in the definition of accidents (e.g., what a serious injury is). Therefore, the classification of an adverse event as accident or incident might vary across organisations and States, this inevitably affecting how safety performance is measured and claimed. Also, it is interesting that according to Boeing (2015, p.6) the selection of departures as exposure to risk is preferred “...since there is a stronger statistical correlation between accidents and departures than there is between accidents

and flight hours, or between accidents and the number of airplanes in service, or between accidents and passenger miles or freight miles". This statement echoes a problem in the industry when establishing indicators: instead of putting efforts in the development of meaningful metrics, the respective decisions might be based on metrics that fit statistical distributions. On one hand, this approach can mislead the conclusions reached through the monitoring of such indicators. On the other hand, this phenomenon might be attributed to the fact that the development of safety metrics remains a vague area because respective uniformity and objective criteria are not provided by authorities and standards.

Since the level of harm experienced or not (i.e., potential harm) is an indication of the safety level achieved, occurrences that have not led to visible losses are actually indications of erosion of safety margins and should be included in safety outcome metrics. Also, due to the sparse number of accidents, the indiscriminate definition and classification of occurrences, the fact that hazards do not always lead to losses and the need to consider interconnectivity of socio-technical systems renders the exclusive use of existing safety outcome metrics insufficient for monitoring safety performance. Thus, safety process metrics are required to complement safety outcome metrics, but currently there hasn't been empirical evidence of how respective proxies relate to safety outcomes within the aviation sector. From a safety model perspective, although latent factors depicted by epidemiological models might serve as proxies, those might be enriched if a systemic model of safety is engaged. Nevertheless, the need for safety metrics, both outcome and process ones, has become more paramount with the introduction of performance-based safety management in aviation.

Perhaps if one looks at quality control more evidence could be found as a link between process indicators and outcome metrics and that is an interesting view since safety management is commonly known to be developed from quality management. Both systems look to control the process to create favourable outcomes. However, there are some important differences, e.g., it is probably easier to find non-quality compared to unsafe situations or occurrences. In quality control there is a focus on a certain product and the upper and lower limits are very well known, e.g., for a jack-screw in an airplane the exact margins are known. For controlling safety outcomes, however, that is not the case. First of all, there are very few companies that produce safety as a product, they transport passengers and goods from one location to another, although safety is a very important requirement to continue operations. When producing those operations only one of the margins is clear, it is relatively easy to predict the economic consequences of delay, but at the same time it really hard to determine when operations become unsafe. Therefore, the analogy of quality control and outcomes does not suit the dynamics of daily operations in aviation.

This review also revealed that many synonyms are available for classifying SPI, the terms 'leading' and 'lagging' being widely used. It is interesting that Øien et al. (2011a) viewed both risk and safety indicators as leading ones, showing that the distinction between leading and lagging metrics might be unclear and misleading. Interestingly, a systems approach was evident in Erikson's classification (Erikson, 2009), who recognised that the terms of leading and lagging make sense only locally, since what comprises an outcome of one process might be input to another. Thus, in the scope of our on-going research, it seems that the use of the terms safety outcome metrics and safety process metrics is more suitable, because the former metrics illustrate what level of safety was achieved, whilst the latter ones are related to how safety has been achieved.

Especially regarding safety culture assessments, the researchers are going to consider them as process indicators; however, we might revise this position in the course of the project.

Moreover, during this review we identified a plethora of safety metrics proposed by the academia and international or regional agencies and authorities, or/and applied by the industry. The initial unfiltered list included more than 600 metrics categorised in ones referring to documented data analysis and metrics that requiring the collection of data through surveys, the latter related mainly to assessment of safety culture characteristics. Following the exclusion of identical and overlapping metrics of the first category, about 160 metrics based on raw/hard data remained in the list; the safety culture assessments were included in one category, due to the high diversity of respective approaches and instruments. In addition, due to the large numbers of the metrics which are based on documented data, we categorised them based on the area of measurement. The areas and methods of measurement we concluded are presented in the Appendix, classified into safety process metrics (Annex 1) and safety outcome metrics (Annex 2). Furthermore, Annex 3 contains a list of safety indicators according to the guidance of SSM and Annex 4 covers process indicators for occurrences per type of operator. Interestingly, the lists included in the Appendix indicate that the types of safety process metrics outnumber the safety outcome ones; however, the vast majority of the published aviation safety statistics focus on the latter as measurements of safety performance.

4 Conclusions

The paper set out to find how safety should be measured according to the standards and recent literature. Instead of finding clear-cut answers, various discussions were found, starting with the definition of safety. Traditionally safety is defined by its absence, which in turn makes it hard to measure, since it can only be recognised when an unsafe situation occurs. Two groups of safety metrics could be identified, process and outcome metrics. Safety process metrics are linked with operational, organisational and safety management activities. Outcomes are occurrences of any severity, problematic is that the thresholds of the types of occurrences are not clearly defined, which makes it hard to compare numbers across industries. Furthermore, accidents and incidents occur very rarely, which makes it hard to use those outcomes to say something about the current level of safety.

The guidance on SPI was found to be limited, companies need to develop their own safety indicators, which creates flexibility and opportunities, but at the same time a common baseline is not established. Although, the standards mention a shift towards a performance-based evaluation of safety, there seem not to be many tools yet to help with that transition.

Taking into account that published aviation safety statistics and industry practice refer mainly to safety outcome indicators, valid safety process indicators which correspond to relevant proxies are required to provide a complete set of safety metrics. However, it seems that it has been difficult to establish valid links between proxies and safety outcomes; so far, there has been little empirical evidence on this topic within aviation and the credibility of safety process metrics depends on the models adopted and reasoning applied. In every case, though, the quality criteria referred in the literature shall be fulfilled when developing safety performance metrics.

The results of this paper are limited by the aviation specific topic and guidance. Furthermore, when discussing metrics, companies often use tools that not always show up directly when searching on the topic of measuring safety, whilst those tools may actually give an indication of the level of safety. The companies themselves are able to point the tools they use themselves, this does not mean that these show up in the academic literature or the guidance given in the aviation industry.

In this review we laid the foundation for the four-year project *Measuring Safety in Aviation – Developing Metrics for Safety Management Systems* and we re-justified the scope of the research: to identify manageable safety metrics that do not require the collection and processing of large amounts of data and provide to the aviation industry valid safety process indicators. In the next step of this project the research team will conduct onsite surveys, will explore why and how partner companies use their safety metrics, will collect relevant data and generate a list of associated safety indicators and evaluate those according to the findings of this review.

Acknowledgements

The research team would like to express their deep thanks to the following members of the knowledge experts group of the project, who reviewed the draft version of this report and provided enlightening and valuable feedback (in alphabetical ascending order of partner organisation):

- Kindunos: John Stoop.
- KLM Cityhopper: Ewout Hiltermann.
- KLu/MLA: Ruud Van Maurik.
- Team-HF: Gesine Hofinger.
- TU Delft: Alexei Sharpanskykh and Frank Guldenmund.

References

- Ale, B.J.M. (2005) 'Tolerable or acceptable: a comparison of risk regulation in the United Kingdom and in the Netherlands', *Risk Analysis*, Vol. 25, No. 2, DOI: 10.1111/j.1539-6924.2005.00585.x
- ARMS Working Group (2010) *The Arms Methodology For Operational Risk Assessment In Aviation Organisations*, Version 4.1, March.
- Aviation Academy (2014) *Project Plan RAAK PRO: Measuring Safety in Aviation – Developing Metrics for Safety Management Systems*, Hogeschool van Amsterdam, Aviation Academy, The Netherlands.
- Bödecker, H. (2013) 'Lufthansa technik group; measurement and driving of safety performance', *Presentation at SM ICG Industry Day*, The Hague, Netherlands, 19 April.
- Boeing (2016) *Statistical Summary of Commercial Jet Planes Accidents: Worldwide Operations 1959–2015*, Aviation Safety Boeing Commercial Airplanes, Seattle [online] http://www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf (accessed 29 September 2016).
- Boishu, Y. (2014) 'SMS and risk assessment automation', *Presentation at SM ICG Industry Day*, Bern, Switzerland, 16 May.

- Bourne, M., Pavlov, A., Franco-Santos, M., Lucianetti, L. and Mura, M. (2013) 'Generating organisational performance', *Int. J. of Op. and Prod Mngemnt.*, Vol. 33, Nos. 11/12, pp.1599–1622, DOI: 10.1108/ijopm-07-2010-0200.
- Cilliers, P. (1998) *Complexity and Postmodernism: Understanding Complex Systems*, Routledge, New York.
- Dekker, S. (2011) *Drift into Failure: From Hunting Broken Parts to Understanding Complex Systems*, Ashgate Publishing Co, Farnham.
- Dekker, S., Cilliers, P. and Hofmeyr, J-H. (2011) 'The complexity of failure: implications of complexity theory for safety investigations', *Safety Science*, Vol. 49, No. 6, pp.939–945.
- Dekker, S.W.A. (2014) *The Field Guide to Understanding Human Error*, 3rd ed., Ashgate Publishing Limited, Farnham.
- Di Gravio, G., Mancini, M., Patriarca, R. and Costantino, F. (2015) 'Overall safety performance of air traffic management system: forecasting and monitoring', *Safety Science*, Vol. 72, pp.351–362, DOI: 10.1016/j.ssci.2014.10.003.
- Duijm, N.J. (2015) 'Recommendations on the use and design of risk matrices', *Safety Science*, Vol. 76, pp.21–31.
- EASA (2014) *A Harmonised European Approach to a Performance Based Environment*, EASA, Cologne.
- EC (2010) 'Regulation (EU) No 996/2010 of the European Parliament and of the Council', *Official Journal of the European Union* [online] <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:295:0035:0050:EN:PDF> (accessed 2 October 2015).
- EC (2014) 'Regulation (EU) No 376/2014 of the European Parliament and of the Council', *Official Journal of the European Union* [online] <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0376&from=EN> (accessed 14 October 2015).
- Erikson, S.G. (2009) 'Performance indicators', *Safety Science*, Vol. 47, No. 4, p.468, DOI: 10.1016/j.ssci.2008.07.024.
- Espig, S. (2013) 'CANSO global approach to safety performance measurement', *Presentation at SMICG Industry Day*, The Hague, Netherlands, 19 April.
- EUROCONTROL (2009) *ATM Safety Framework Maturity Survey, Methodology for ANSPs* [online] <http://skybrary.aero/bookshelf/books/1256.pdf> (accessed 6 November 2015).
- FAA (2013) *Safety Management System, Order 8000.369A*, Federal Aviation Administration, Washington [online] <http://www.faa.gov/documentLibrary/media/Order/8000.369A.pdf> (accessed 2 October 2015).
- Flight Global (2016) *Airline Safety and Losses – Annual Review 2015* [online] <https://flightglobal.com/asset/6729> (accessed 12 September 2016).
- Grabowski, M., Ayyalasomayajula, P., Merrick, J., Harrald, J.R. and Roberts, K. (2007) 'Leading indicators of safety in virtual organizations', *Safety Science*, Vol. 45, No. 10, pp.1013–1043, DOI: 10.1016/j.ssci.2006.09.007
- Grote, G. (2012) 'Safety management in different high-risk domains – all the same?', *Safety Science*, Vol. 50, No. 10, pp.1983–1992, DOI: 10.1016/j.ssci.2011.07.017.
- Guldenmund, F.W. (2007) 'The use of questionnaires in safety culture research – an evaluation', *Safety Science*, Vol. 45, No. 6, pp.723–743.
- Hale, A. (2009) 'Why safety performance indicators?', *Safety Science*, Vol. 47, No. 4, pp.479–480, DOI: 10.1016/j.ssci.2008.07.018.
- Hale, A., Guldenmund, F., Loenhout, P.V. and Oh, J. (2010) 'Evaluating safety management and culture interventions to improve safety: effective intervention strategies', *Safety Science*, Vol. 48, pp.1026–1035.

- Harms-Ringdahl, L. (2009) 'Dimensions in safety indicators', *Safety Science*, Vol. 47, No. 4, pp.481–482, DOI: 10.1016/j.ssci.2008.07.019.
- Hinze, J., Thurman, S. and Wehle, A. (2013) 'Leading indicators of construction safety performance', *Safety Science*, Vol. 51, No. 1, pp.23–28, DOI: 10.1016/j.ssci.2012.05.016.
- Hollnagel, E. (2010) *On How (Not) to Learn from Accidents* [online] http://www.uis.no/getfile.php/Konferanser/Presentasjoner/Ulykkesgransking%202010/EH_AccLearn_short.pdf (accessed 15 November 2015).
- Hollnagel, E. (2012) *The Health Foundation Inspiring Improvement: Proactive approaches to safety management*, Pamphlet, The Health Foundation [online] <http://www.health.org.uk/sites/default/files/ProactiveApproachesToSafetyManagement.pdf> (accessed 15 November 2015).
- Hollnagel, E. (2014) *Safety-I and Safety-II: The Past and Future of Safety Management*, Ashgate Publishing, Ltd.
- Holt, C. (2014) 'Safety intelligence and management workshop', *Presentation at the Safety Intelligence and Management Workshop*, Dubai, UAE, November.
- Hopkins, A. (2012) *Disastrous Decisions: The Human and Organisational Causes of the Gulf of Mexico Blowout*, CCH Australia Ltd, Australia.
- Hubbard, D. and Evans, D. (2010) 'Problems with scoring methods and ordinal scales in risk assessment', *IBM Journal of Research and Development*, Vol. 54, No. 3, pp.2:1–2:10.
- IATA (2016) *Safety Report 2015*, International Air Transport Association. Montreal, Geneva [online] <http://www.iata.org/publications/Documents/iata-safety-report-2015.pdf> (accessed 12 September 2016).
- ICAO (2010) *Annex 13 – Aircraft Accident and Incident Investigation*, 10th ed., International Civil Aviation Organization. Montréal, Canada.
- ICAO (2013a) *Annex 19 – Safety Management*, 1st ed., International Civil Aviation Organization. Montréal, Canada.
- ICAO (2013b) *Doc 9859, Safety Management Manual (SMM)*, 3rd ed., International Civil Aviation Organization. Montréal, Canada.
- ICAO (2016) *ICAO Safety Report 2016 Edition*, International Civil Aviation Organization. Montréal, Canada.
- ISO (1999) *Safety Aspects – Guidelines for Their Inclusion in Standards*, ISO/IEC guide 51:1999, International Organisation for Standardisation, Geneva, Switzerland.
- Karanikas, N. (2015a) 'Correlation of changes in the employment costs and average task load with rates of accidents attributed to human error', *Aviation Psychology and Applied Human Factors*, Vol. 5, No. 2, pp.104–113, DOI: 10.1027/2192-0923/a000083.
- Karanikas, N. (2015b) 'An introduction of accidents' classification based on their outcome control', *Safety Science*, Vol. 72, pp.182–189, DOI: 10.1016/j.ssci.2014.09.006.
- Karanikas, N. (2016) 'Critical review of safety performance metrics', *International Journal of Business Performance Management*, Vol. 17, No. 3, pp.266–285.
- Kjellen, U. (2000) *Prevention of Accidents Through Experience Feedback*, Taylor and Francis, London.
- Kruijssen, E. (2013) 'Deriving safety metrics, from data to intelligence', *Presentation at Industry Day*, The Hague, Netherlands, 19 April.
- Lardner, R., McCormick, P. and Novatsis, E. (2011) 'Testing the validity and reliability of a safety culture model using process and occupational safety performance data', Paper Presented at *Hazards XXII – Process Safety and Environmental Protection Conference*, Liverpool, UK, 11–14 April.

- Leveson, N. (2011) *Engineering a Safer World: Systems Thinking Applied to Safety*, MIT Press, Boston, Mass.
- Lofquist, E.A. (2010) 'The art of measuring nothing: the paradox of measuring safety in a changing civil aviation industry using traditional safety metrics', *Safety Science*, Vol. 48, pp.1520–1529, DOI: 10.1016/j.ssci.2010.05.006.
- Maurino, D. (2005) *Threat and Error Management (TEM) (PDF), Coordinator, Flight safety and Human Factors Programme – ICAO, Canadian Aviation Safety Seminar (CASS)*.
- O'Connor, P., O'Dea, A., Kennedy, Q. and Buttrey, S.E. (2011) 'Measuring safety climate in aviation: a review and recommendations for the future', *Safety Science*, Vol. 49, No. 2, pp.128–138, DOI: 10.1016/j.ssci.2010.10.001.
- Øien, K., Utne, I.B. and Herrera, I.A. (2011a) 'Building safety indicators: part 1 – theoretical foundation', *Safety Science*, Vol. 49, No. 2, pp.148–161, DOI: 10.1016/j.ssci.2010.05.012.
- Øien, K., Utne, I.B., Tinmannsvik, R.K. and Massaiu, S. (2011b) 'Building safety indicators: part 2 – application, practices and results', *Safety Science*, Vol. 49, No. 2, pp.162–171, DOI: 10.1016/j.ssci.2010.05.015.
- Papadimitriou, E., Yannis, G., Bijleveld, F. and Cardoso, J.L. (2013) 'Exposure data and risk indicators for safety performance assessment in Europe', *Accident Analysis and Prevention*, Vol. 60, pp.371–383, DOI: 10.1016/j.aap.2013.04.040.
- Pasman, H. and Rogers, W. (2014) 'How can we use the information provided by process safety performance indicators? Possibilities and limitations', *Journal of Loss Prevention in the Process Industries*, Vol. 30, pp.197–206, DOI: 10.1016/j.jlp.2013.06.001.
- Perrin, E. (2014) 'Advancing thinking on safety performance indicators', *Presentation at the Safety Intelligence and Management Workshop*, Dubai, UAE, November.
- Piric, S. (2011) *The Relation between Safety Culture and Organisational Commitment: Differences between Low-Cost and Full-Service Carrier Pilots*, MSc Thesis, Cranfield University, UK.
- Podgórski, D. (2015) 'Measuring operational performance of OSH management system – a demonstration of AHP-based selection of leading key performance indicators', *Safety Science*, Vol. 73, pp.146–166, DOI: 10.1016/j.ssci.2014.11.018.
- Rasmussen, J. (1997) 'Risk management in a dynamic society: a modelling problem', *Safety Science*, Vol. 27, Nos. 2–3, pp.183–213.
- Reason, J. (1990) *Human Error*, Cambridge University Press, New York.
- Reiman, T. and Pietikäinen, E. (2012) 'Leading indicators of system safety – monitoring and driving the organizational safety potential', *Safety Science*, Vol. 50, No. 10, pp.1993–2000, DOI: 10.1016/j.ssci.2011.07.015.
- Rockwell, T.H. (1959) 'Safety performance measurement', *Journal of Industrial Engineering*, Vol. 10, No. 1, pp.12–16.
- Roelen, A.L.C. and Klompstra, M.B. (2012) 'The challenges in defining aviation safety performance indicators', *PSAM 11 and ESREL 2012*, 25–29 June, Helsinki, Finland.
- Safety Management International Collaboration Group (SMICG, 2013) *Measuring Safety Performance Guidelines for Service Providers* [online] <http://www.skybrary.aero/bookshelf/books/2395.pdf> (accessed 15 September 2015).
- Safety Management International Collaboration Group (SMICG, 2014). *A Systems Approach to Measuring Safety Performance: A Regulator Perspective* [online] <http://live.transport.gov.mt/admin/uploads/media-library/files/A%20Systems%20Approach%20to%20Measuring%20Safety%20Performance%20-%20the%20regulator%20perspective.pdf>.

- Salmon, P., McClure, R. and Stanton, N. (2012) 'Road transport in drift? Applying contemporary systems thinking to road safety', *Safety Science*, Vol. 50, No. 9, pp.1829–1838.
- Saracino, A., Antonioni, G., Spadoni, G., Guglielmi, D., Dottori, E., Flamigni, L., Pacini, V. (2015) 'Quantitative assessment of occupational safety and health: application of a general methodology to an Italian multi-utility company', *Safety Science*, Vol. 72, pp.75–82, DOI: 10.1016/j.ssci.2014.08.007.
- SIA (2015) *Besluit inzake aanvraag subsidie regeling RAAK-PRO 2014 voor het project Measuring Safety in Aviation – Developing Metrics for Safety Management Systems*, (projectnummer:2014-01-11ePRO) Kenmerk: 2015-456, Nationaal Regieorgaan Praktijkgericht Onderzoek SIA, The Netherlands.
- Sinelnikov, S., Inouye, J. and Kerper, S. (2015) 'Using leading indicators to measure occupational health and safety performance', *Safety Science*, Vol. 72, pp.240–248, DOI: 10.1016/j.ssci.2014.09.010.
- Singh, S., Darwish, T.K., Costa, A.C. and Anderson, N. (2012) 'Measuring HRM and organisational performance: concepts, issues and framework', *Management Decision*, Vol. 50, No. 4, pp.651–667, retrieved from Google Scholar.
- Snowden, D.J. and Boone, M.E. (2007) 'A leader's framework for decision making', *Harvard Business Review*, Vol. 85, No. 11, p.68.
- Stolzer, A.J., Halford, C.D. and Goglia, J.J. (2008) *Safety Management Systems in Aviation*, Ashgate, Aldershot, Hampshire, England, Burlington, VT, retrieved from Library of Congress or OCLC Worldcat.
- Tump, R. (2014) 'NLR flight test and SMS; what to measure?', *Presentation at the European Flight Test Safety Workshop 2014*, Manching, Germany, 5 November.
- UK CAA (2011) *Safety Plan 2011 to 2013*, Civil Aviation Authority, London, UK
- Underwood, P. and Waterson, P. (2013) *Accident Analysis Models and Methods: Guidance for Safety Professionals*.
- Wahlström, B. and Rollenhagen, C. (2014) 'Safety management – a multi-level control problem', *Safety Science*, Vol. 69, pp.3–17, DOI: 10.1016/j.ssci.2013.06.002.
- Webb, P. (2009) 'Process safety performance indicators: a contribution to the debate', *Safety Science*, Vol. 47, No. 4, pp.502–507, DOI: 10.1016/j.ssci.2008.07.029.
- Weick, K.E. and Sutcliffe, K.M. (2001) *Managing the Unexpected: Assuring High Performance in an Age of Complexity*, 1 ed., JOSSEY-BASS, San Francisco.
- Woods, D.D. (2006) 'Essential characteristics of resilience', in *Resilience Engineering: Concepts and Precepts*, Ashgate, Aldershot.
- Woods, D.D., Branlat, M., Herrera, I. and Woltjer, R. (2015) 'Where is the organization looking in order to be proactive about safety? A framework for revealing whether it is mostly looking back, also looking forward or simply looking away', *Journal of Contingencies and Crisis Management*, Vol. 23, No. 2, pp.97–105, DOI: 10.1111/1468-5973.12079.
- Wreathall, J. (2009). Leading? Lagging? Whatever!, *Safety Science*, Vol. 47, No. 4, pp.493–494, DOI: 10.1016/j.ssci.2008.07.031.

Appendix

Table A1 Safety process metrics

<i>Areas of measurement</i>	<i>Methods of measurement</i>
Compliance	% of compliance or non-compliance against a list of topics
Productivity of safety management activities over time	Number of open or closed safety issues (e.g., number of risk mitigation measures pending, number of hazard reports not followed-up)
	% of open or closed safety issues
	Average time to respond to safety issues
	Average time to close safety issues
	Average time of open safety issues
	Number of activities (e.g., number of audits conducted)
	% of accomplished or non-accomplished activities
	Average exposure of target population to each safety management activity (e.g., safety training hours per employee, safety audits per area)
	% of target population covered by each activity (e.g., % of staff and managers trained in (specific) safety topics, % of working activities covered in safety audits, % of reported accidents and incidents investigated, % of working instructions revised as result of risk mitigation measures)
	Ratio of time allotted to (specific) activities / overall time (e.g., hours in safety training / total training hours)
Frequency of activities	
Effectiveness of safety management activities over time	% of objectives and targets met (e.g., % of workers assessed competent after safety training, % of risks decreased)
Contribution to safety management activities over time	Number of contributions (e.g., number of managers attending safety meetings, hazard reports submitted)
	Ratios of contributions / population (e.g., number of contractors submitted hazard reports / total number of contractors)
Best practice in safety management over time	Number of activities following best practice (e.g., number of meetings dedicated to human performance, number of safety conferences attended, number of safety bulletins published, number of new safety goals and objectives, number of safety performance indicators subject to external benchmarking)
	% of activities following best practice (e.g., % of leading indicators, % risks controls relied on operators' performance)
Human resources for safety management activities	Number of staff carrying out activities (e.g., number of qualified accident investigators)
	Ratios of available / required staff to support activities (e.g., safety officers available / safety officers foreseen)
Human resources over time	Number of staff with specific competencies (e.g., pilots with a current licence)
	Staff turn-over
Exposure to risk	Number of known exemptions (e.g., number of technical modifications pending)
	Probability of unsafe events
	Number of unexpected disturbances (e.g., flights delayed, flight plans changed)
	Duration of unexpected disturbances (e.g., time delays)
	Ratio of known exemptions/activity unit (e.g., Minimum Equipment List entries / flight hours)
Culture	Number of risks per level (e.g., low risks)
	% of specific risk levels (e.g., % of medium risks)
	Surveys (e.g., extent to which safety culture characteristics are present)
	Analysis of records (e.g., % of workers participating in safety review meetings)
Financial	Comparisons (e.g., changes of safety culture characteristics over time)
	Ratio of budget invested on (specific) safety management activities / quantified costs of events led to losses

Table A2 Safety outcome metrics

<i>Areas of measurement</i>	<i>Methods of measurement</i>
Accidents and occurrences over a given period	Number of events (e.g., accidents, serious incidents, near-misses) Ratio of events/activity unit (e.g., accidents/flight departures, occurrences/year)
Losses over a given period	Number of losses (e.g., fatalities, injuries, hull losses) Ratio of losses/activity unit (e.g., fatalities/passenger miles flown) Time lost due to losses (e.g., lost time due to injuries) Quantified costs of losses

Table A3 Safety indicators adapted from SMM

<i>Air operators</i>	
<i>Lower-consequence indicators (event/activity-based)</i>	<i>High-consequence indicators (occurrence/outcome-based)</i>
Operator combined fleet monthly incident rate (e.g., per 1,000 FH)	Air operator individual fleet monthly serious incident rate (e.g., per 1 000 FH)
Operator internal QMS/SMS annual audit LEI % or findings rate (findings per audit)	Air operator combined fleet monthly serious incident rate (e.g., per 1 000 FH)
Operator voluntary hazard report rate (e.g., per 1,000 FH)	Air operator engine IFSD incident rate (e.g., per 1,000 FH)
Operator DGR incident report rate (e.g., per 1 000 FH)	
<i>Aerodrome operators</i>	
Aerodrome operator internal QMS/SMS annual audit LEI % or findings rate (findings per audit)	Aerodrome operator quarterly ground accident/serious incident rate — involving any aircraft (e.g., per 10 000 ground movements)
Aerodrome operator quarterly runway foreign object/debris hazard report rate (e.g., per 10 000 ground movements)	Aerodrome operator quarterly runway excursion incident rate — involving any aircraft (e.g., per 10 000 departures)
Operator voluntary hazard report rate (per operational personnel per quarter)	Aerodrome operator quarterly runway incursion incident rate — involving any aircraft (e.g., per 10 000 departures)
Aerodrome operator quarterly aircraft ground foreign object damage incident report rate — involving damage to aircraft (e.g., per 10 000 ground movements)	
<i>ATS operators</i>	
ATS operator quarterly FIR TCAS RA incident rate — involving any aircraft (e.g., per 100.000 flight movements)	ATS operator quarterly FIR serious incident rate — involving any aircraft (e.g., per 100 000 flight movements)
ATS operator quarterly FIR level bust (LOS) incident rate — involving any aircraft (e.g., per 100 000 flight movements)	ATS operator quarterly/annual near-miss incident rate (e.g., per 100 000 flight movements)
ATS operator internal QMS/SMS annual audit LEI % or findings rate (findings per audit)	

Source: ICAO (2013b)

Table A3 Safety indicators adapted from SMM (continued)

<i>POA/DOA/MRO ORGANIZATIONS</i>	
MRO/POA/DOA internal QMS/SMS annual audit LEI % or findings rate (findings per audit)	MRO/POA quarterly rate of component technical warranty claims
MRO/POA/DOA quarterly final inspection/testing failure/rejection rate (due to internal quality issues)	POA/DOA quarterly rate of operational products which are the subject of ADs/ASBs (per product line)
MRO/POA/DOA voluntary hazard report rate (per operational personnel per quarter)	MRO/POA quarterly rate of component mandatory/major defect reports raised (due to internal quality issues)

Source: ICAO (2013b)

Table 4 Safety indicators per area of operation

<i>Operators</i>	<i>Occurrence</i>	<i>Indicator</i>	
Air operators	Traffic collision	Number of traffic collision avoidance system (TCAS) resolution advisories per 1,000 flight hours (FH)	
	Runway excursion	Number of unstabilised approaches per 1,000 landings	
	Ground collision	Number of runway incursions per 1,000 take-offs	
	Controlled flight into terrain	Number of ground proximity warning system (GPWS) and enhanced ground proximity warning system (EGPWS) warnings per 100 take-offs	
	Accident/incident related to poor flight preparation		Number of cases where flight preparation had to be done in less than the normally allocated time
			Number of short fuel events per 100 flights
	Accident/incident related to fatigue		Number of fuel calculation errors per 100 flights
			Number of extensions to flight duty periods per month/quarter/year and trends
	Accident/incident related to ground-handling		Number of incidents with ground handlers per month/quarter/year and trends
			Number of dysfunctions per ground handler per month/quarter/year and trends
			Number of mass and balance errors per ground handler per month/quarter/year and trends
	Maintenance related accident/incidents		Pilots reports (PIREPS) per 100 take offs
			Deferred items per month and aircraft
			In flight shut down (IFSD) per 1,000 FH
		In flight turn backs (IFTB) and deviations per 100 take offs	
		Number of service difficulty reports filed with the Civil Aviation Authority	
	Number of delays of more than 15 minutes due to technical issues per 100 take offs		
	Number of cancellations per 100 scheduled flights due to technical issues		
	Rejected take offs per 100 take-offs due to technical issues		

Source: Adapted from SMICG (2013)

Table 4 Safety indicators per area of operation (continued)

<i>Operators</i>	<i>Occurrence</i>	<i>Indicator</i>	
ANSP	Traffic collision	Number of level busts/exposure	
		Number of TCAS required action (RA) (with and without loss of separation)/exposure	
		Number of minimum separation infringement/exposure	
		Number of inappropriate separation (airspace in which separation minima is not applicable)/exposure	
		Number of aircraft deviation from air traffic control (ATC) clearance/exposure	
		Number of airspace infringements/exposures	
		Number of aircraft deviations from air traffic management (ATM) procedures/exposure	
		Number of inappropriate or absences of ATC assistance to aircraft in distress	
		Number of near Controlled Flight Into Terrain (CFIT) IFSD/exposure	
		Number of inappropriate ATC instruction (no instruction, wrong information, action communicated too late, etc.)	
Airports	Traffic collision/controlled flight into terrain	% of runway incursions where no avoiding action was necessary	
		% of runway incursion where avoiding action was necessary	
Airports	controlled flight into terrain	Number of near Controlled Flight Into Terrain (CFIT) IFSD/exposure	
		Number of inappropriate ATC instruction (no instruction, wrong information, action communicated too late, etc.)	
		% of runway incursions where no avoiding action was necessary	
	Runway excursion	Runway incursion	% of runway incursion where avoiding action was necessary
			Fire extinguishing services (ICAO airport fire fighting categories) decrease in value (# decrease- hours/ # airport annual operating hours)
	Runway incursion	Runway incursion	Number of radio/phone failures per 100 operations
			Number of fire rescue vehicles failures per 100 operations
			Runway incursions per 1,000 operations
			Number of failures or defects found during routine inspection (signage)
	Collision with vehicle on ground/ground-equipment	Ground collision with wildlife	Number of defects reported (signage)
Average lead-time for repair/replacement (signage)			
Notified platform safety rules violations per 1,000 operations			
Foreign object damage (FOD)	Runway incursion	Number of ground collisions with wildlife	
		Number of inspections of fences and other protective devices per month/quarter/year	
Runway incursion	Runway incursion	Number of FOD found during routine inspections	
		Number of FOD found out of inspections and after report	
		Number of failures or defects found during routine inspection (lights)	
Bird-strike in flight shut down (IFSD)	Bird-strike in flight shut down (IFSD)	Number of defects reported (lights)	
		average lead-time for repair/replacement (per month/quarter/year and trends) (lights)	
		Number IFSD per 10,000 FH following bird-strike	

Source: Adapted from SMICG (2013)