

RAAK PRO Project: Measuring Safety in Aviation

Deliverable: Review of Existing Aviation Safety Metrics

Authors: Steffen Kaspers, Nektarios Karanikas, Alfred Roelen, Selma Piric, Robert J. de Boer
Project number: S10931

RAAK PRO Project: Measuring Safety in Aviation

Review of Existing Aviation Safety Metrics

Steffen Kaspers¹, Nektarios Karanikas¹, Alfred Roelen^{1,2},
Selma Piric¹, Robert J. de Boer¹

¹Aviation Academy, Amsterdam University of Applied Sciences, the Netherlands

²NLR, Amsterdam, the Netherlands

Contents

1. INTRODUCTION	3
2. REVIEW OF LITERATURE AND INDUSTRY REFERENCES	4
2.1 Views on Safety.....	4
2.2 Safety performance metrics.....	5
2.3 Classification of safety performance metrics	6
2.4 Safety outcome metrics	7
2.5 Safety process metrics	7
2.6 Quality of metrics.....	8
3. DISCUSSION	8
4. CONCLUSIONS	9
ACKNOWLEDGMENTS	10
REFERENCES	10
APPENDIX	15
Annex 1: Safety process metrics	15
Annex 2: Safety outcome metrics	17

1. Introduction

The improvement of aviation safety has been a focal point for companies and authorities. Although statistics show low rates of accidents (e.g., ICAO, 2015), 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 emphasizing 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 will render safety management more proactive than it currently is; the proper monitoring of the respective safety management activities and the outcomes of those will allow identifying and managing flaws before accidents occur.

However, proactive safety relies heavily on the use of relevant data sourcing from day-to-day activities and operations; the concept is that such data will be processed in a way that will allow to timely identify and control new hazards and combinations of hazards, the deviations from standards included. Under this concept, it seems that although large companies might collect large amount of safety-related data and establish proactive safety metrics, small and medium enterprises (SMEs) lack high volumes of 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, Branlat, Herrera, & Woltjer, 2015; Lofquist, 2010) and considerable amount of resources are required to process high volumes of data (e.g., Kruijssen, 2013).

Following relevant concerns of Dutch aviation companies, the Aviation Academy of the Amsterdam University of Applied Sciences initiated a research project entitled “Measuring Safety in Aviation – Developing Metrics for Safety Management Systems”. The aim of the project is to identify ways to measure operational safety without the benefit of large amounts of safety data (Aviation Academy, 2014). The researchers will examine the validity of current safety metrics, explore new suitable safety metrics based on existing and alternative models and approaches to safety, generate and validate a short list of suitable safety metrics, and translate this knowledge into a web-based dashboard for the industry. The project will last four years, from September 2015 to September 2019, 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 industry, academia and authorities’ representatives.

As a first step in this project, this review charts the current suggestions and practice in safety metrics and will comprise the foundation for the next project phases. 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 Scencedirect 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). In addition to the necessity of this report for moving to the next steps of the project, we envisage that it will be a self-standing useful reference for aviation professionals and safety scientists.

Although the scope of the review is about safety metrics, the latter cannot be viewed outside their context. Hence, the report starts with presenting various views on safety and the challenges when measuring safety. Next, we review the references about safety metrics, the role of Safety Management Systems (SMS) in safety monitoring, the classifications of safety performance indicators (SPI), and the quality criteria of metrics in general. The report continues with a discussion on the literature and industry references reviewed and concludes with the connection of this report with the next phase of the project, where surveys will allow us to review the application of safety metrics in practice.

2. Review of Literature and Industry References

2.1 Views on Safety

Long-established views on safety and relevant limitations

The International Standardization Organization (ISO) defines safety as “...freedom from unacceptable risk...”; risk “...is a combination of the probability of occurrence of harm and the severity of the harm...”, and “...harm is 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, p. 2-1). Both definitions include the term risk, which is defined as a combination of probability and severity of harm, and refer to acceptable levels of risk, thus suggesting the existence of a threshold that distinguishes 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 et al., 2010). Likewise, the severity of harm is generated 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 may be true for general categories of events (e.g., controlled flights into terrain, runway excursions) but not always feasible when considering combinations of various factors that can contribute in those high-level events (Roelen 2012; Leveson 2011).

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 organizations to define their own risk tolerances and thresholds, rendering thus actually cumbersome to make respective comparisons across the aviation industry. Furthermore, the acceptability of risk depends on the system considered; for example, 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, Yannis, Bijleveld, & Cardoso, 2013; Pasmaan & Rogers 2014; Sinelnikov et al., 2015).

Ale (2005) suggested a maximum acceptable individual fatality risk of 1×10^{-6} per year in the Netherlands, and identified a strong sensitivity of the public to multiple fatalities resulting from a single event. 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 so that actions to prevent reoccurrence become unavoidable, 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 organization 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 accidents attributed to human errors (Karanikas, 2015a).

Alternative views on safety

Weick & Sutcliffe (2001, p. 30) defined safety as “...a dynamic non-event...”. The authors stress that we recognise safety by the absence of harm (i.e. something bad not happening) in a constantly changing context, so actually we define safety through non-safety. Various authors (e.g., Dekker, Cilliers, & Hofmeyr, 2011; Cilliers, 1998; Dekker, 2011 cited in Salmon, McClure, & Stanton, 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 was not fully understood before the event (Snowden & Boone, 2007); Lofquist (2010) addressed the need to consider the interactivity within socio-technical systems when measuring safety.

Hollnagel (2014) introduced the concept of Safety -II, where safety is defined as a “...system’s ability to succeed under varying conditions, so that the number of intended and acceptable outcomes is as high as possible...”. The aforementioned author stresses 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 & Rollenbogen, 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 safety performance indicators in the frame of an SMS (UK CAA, 2011; Holt, 2014). According to ICAO (2013a, p.1-2) safety performance is “A State or a service provider’s safety achievement as defined by its safety performance targets and safety performance indicators”, “[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 safety performance indicators as part of its SMS. Within the SMS, another distinction is made: high consequence indicators, which are accident and serious incident based (e.g., air operator monthly serious incident rate of an operator’s individual fleet); low consequence indicators, which are based on activities and incidents (e.g., voluntary hazard report rate per operational personnel per quarter). 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. L 122/25). Also, the European Commission (EC, 2014 p. L 122/25) 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; the difference between an accident and a serious incident lies only in the 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 system, tier 2 indicators depict safety management performance of operators and tier 3 metrics address the activities of the regulator (SMICG, 2014). Safety performance indicators 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. Safety performance indicators 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, Guldenmund, Loenhout, & Oh, 2010).

Also, safety management is often linked to safety culture (e.g., Stolzer, Halford, & Goglia, 2008), the latter lacking a common definition in the literature (Guldenmund, 2007) and being assessed with various instruments (Karanikas, in press). The European Union’s Single European Sky Performance Scheme adds the assessment of Just Culture within an organisation as a leading indicator (EUROCONTROL, 2009). However, 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 much of the professional and some scientific literature, safety performance indicators are classified as “lagging” or “leading”. Grabowski, Ayyalasomayajula, Merrick, Harrald, & Roberts (2007, p.1017) state: “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 prioritize safety management activities and determine actions for safety improvement.

Harms-Ringdahl (2009) proposed the use of the terms activity and outcome indicators in correspondence with leading and lagging indicators. Reiman & Pietikäinen, (2012) make a distinction within leading indicators between driving and monitoring ones. 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) suggests that leading indicators are seen as inputs and lagging indicators are viewed as outputs, thus all indicators might be characterized as both leading and lagging depending on their place in the process. Øien et al. (2011) define 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.

In an attempt for a more elaborate classification than simply leading and lagging, Hinze et al., (2013) suggests the distinction of safety leading indicators to passive and active. Passive leading indicators address the state of safety in long term or 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 h of exposure). Hale (2009) addresses the confusion about leading and lagging indicators and attributes this to variances in: (1) the ‘degree’ of leading, (2) compression of the temporal dimension, and (3) the categorisation of causal factors (e.g. unsafe acts, unsafe conditions).

Therefore, as Hinze et al., (2013) recognises, multiple terms are used for leading and lagging indicators. Table 1 shows the various classifications discussed by the authors cited in this section:

Safety Process Metrics	Safety Outcome Metrics
Leading indicators	Lagging indicators
Upstream indicators	Downstream indicators
Predictive indicators	Historical indicators
Heading indicators	Trailing indicators
Positive indicators	Negative indicators
Active indicators	Reactive indicators
Predictive indicators	Retrospective indicators
Input indicators	Output indicators
Driving/monitoring indicators	Lagging indicators
Proactive indicators	Reactive indicators
Activity (Based) indicators	Outcome (Based) indicators

Table 1: Pairs of indicator types referred in literature

2.4 Safety outcome metrics

As ICAO (2013a) and the European Commission (EC, 2014) mention, the reporting of occurrences (i.e. incidents and serious incidents) is primarily aimed 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 organizations (e.g., ICAO, 2015; Flightglobal, 2016; IATA, 2015; Boeing, 2015). 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, Pavlov, Franco-Santos, Lucianetti, & Mura, 2013; Singh, Darwish, Costa, & Anderson, 2012) recognised that association of performance with quality of 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 outcomes (ICAO, 2013b). As Karanikas (2015b) concluded, the mere reference to severity of events without prior considering their potential is not representative of safety performance. Hence, in this report we will use the term safety outcome metrics for all metrics referring to accidents and occurrences. Some recent proposals for safety performance metrics and assessment methods based on outcomes (i.e. accidents and occurrences) are the following:

- Di Gravio, Manchini, Patriarca, & Constantino (2015) proposed a statistical analysis of safety events based on Monte Carlo simulation and weighing of factors, as means to develop proactive safety indicators.
- 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 (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 accidents and 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, O'Dea, Kennedy, & Buttrey 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, Utne, Tinmannsvik, & Massaiu, 2011). In this report we refer to those as safety process metrics in order to distinguish them from safety outcome metrics.

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 & Pietikäinen 2012). Therefore, the validity of process metrics 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 & 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. 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 & Error Management (e.g., Maurino, 2005) and Tripod (e.g., Kjellen, 2000).

- Systemic models such as STAMP (Leveson, 2011), FRAM (Hollnagel, 2010) and Accimap (e.g., Rasmussen, 1997) that focus on component interactions rather than single component failures in a dynamic, variable and interactive operational context.

2.6 Quality of metrics

Karanikas (in press) discussed the limited practicality, validity and ethicality of some safety metrics proposed in literature or applied in practice. Various authors mention 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, Thurman, & Wehle, 2013; Karanikas, in press; Podgórski 2015; Sinelnikov, Inouye, & Kerper, 2015; Webb 2009; Øien et al., 2011; Øien, Utne, Tinmannsvik, & Massaiu, 2011; Rockwell, 1959). The following list consolidates the quality criteria the aforementioned authors refer to:

- 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) suggest 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 to (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. Discussion

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 as 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 recognized in industry.

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 organizations and States, this inevitably affecting how safety performance is measured and claimed. Also, it is interesting that according to Boeing (2015) 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” (Boeing, 2015 p. 6). This statement represents 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. This approach can mislead the conclusions reached through the monitoring of such indicators. 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. 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.

This review also revealed that many synonyms are available for classifying safety performance indicators, the terms “leading” and “lagging” being widely used. It is interesting that Øien et al. (2011) 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 thinking 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. In the scope of this research project we are going to use the terms safety outcome metrics and safety process metrics, because the former illustrates what level of safety was achieved, whilst the latter is related to how safety has been achieved. Especially regarding safety culture assessments, at this stage we are going to consider them as process indicators; however, we might revise this position in the course of the project.

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 that referred to documented data analysis, and metrics that required 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 categorized 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). 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. Thus, it is important for the research team to search to what extent existing safety process metrics are used by the industry and the corresponding reasons (e.g., lack or not recording of relevant data, unavailability of resources for monitoring safety process indicators).

4. Conclusions

Taking into account that published aviation safety statistics and industry practice refer mainly to safety outcome indicators, valid safety process indicators, founded on relevant proxies, are required to provide a complete set of safety performance 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 in this topic and the credibility of safety process metrics depends on the models adopted and reasoning applied. In every case though, the fulfilment of the quality criteria referred in the literature will comprise useful references when developing safety performance metrics.

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 on-site 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.

Acknowledgments

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 organization):

- Kindunos: John Stoop
- KLM Cityhopper: Ewout Hiltermann
- KLu / MLA: Ruud Van Maurik
- Team-HF: Gesine Hofinger
- TU Delft: Alexei Sharpanskykh & 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, 2005. DOI: 10.1111/j.1539-6924.2005.00585.x

ARMS Working Group. (2010). The ARMS methodology for operational risk assessment in aviation organisations. V4.1, March 2010

Aviation Academy. (2014) “Project Plan RAAK PRO: Measuring safety in aviation – developing metrics for Safety Management Systems”, Hogeschool van Amsterdam, Aviation Academy, The Netherlands.

Boeing (2015). Statistical Summary of Commercial Jet Planes Accidents: Worldwide Operations 1959-2014. Aviation Safety Boeing Commercial Airplanes, Seattle. Retrieved from http://www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf

Boishu, Y., (2014). SMS and Risk assessment automation, presentation at SM ICG industry day, Bern, Switzerland, 16 May 2014)

Bourne, M., Pavlov, A., Franco-Santos, M., Lucianetti, L., & Mura, M. (2013). Generating organisational performance. *Int Jnl of Op & Prod Mngemnt*, 33(11/12), 1599-1622. doi:10.1108/ijopm-07-2010-0200

Bödecker, H. (2013). Lufthansa Technik Group; Measurement and driving of safety performance. Presentation at SM ICG industry day, The Hague, Netherlands, 19 april 2013.

UK CAA (2011). Safety Plan 2011 to 2013, Civil Aviation Authority, London, UK

Cilliers, P. (1998). Complexity and postmodernism: Understanding complex systems New York: Routledge.

Dekker, S., W., A., (2014) The Field Guide to Understanding Human Error (3rd Ed). Ashgate.

Dekker, S., Cilliers, P., & Hofmeyr, J.-H. (2011). The complexity of failure: Implications of complexity theory for safety investigations. *Safety Science*, 49(6), 939–945.

Di Gravio, G., Mancini, M., Patriarca, R., & Costantino, F. (2015). Overall safety performance of air traffic management system: Forecasting and monitoring. *Safety Science*, 72, 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*, 76, 21-31

- EASA. (2014). A Harmonised European Approach to a Performance Based Environment. Cologne: EASA.
- EC, (2010) REGULATION (EU) No 996/2010 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL (Official Journal of the European Union) Retrieved from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:295:0035:0050:EN:PDF>
- EC, (2014) REGULATION (EU) No 376/2014 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL (Official Journal of the European Union) Retrieved from <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0376&from=EN>
- Erikson, S. G. (2009). Performance indicators. *Safety Science*, 47(4), 468. doi:10.1016/j.ssci.2008.07.024
- Espig, S. (2013) CANSO Global Approach to Safety Performance Measurement, Presentation at SM ICG industry day, The Hague, Netherlands, 19 April 2013.
- EUROCONTROL (2009). ATM Safety Framework Maturity Survey, Methodology for ANSPs. Available at <http://skybrary.aero/bookshelf/books/1256.pdf>
- FAA (2013). Safety Management System, Order 8000.369A, Washington: Federal Aviation Administration. Retrieved from <http://www.faa.gov/documentLibrary/media/Order/8000.369A.pdf>.
- Flight global (2016). Airline Safety & Losses – Annual Review 2015. Retrieved from <https://flightglobal.com/asset/6729>
- Grabowski, M., Ayyalasomayajula, P., Merrick, J., Harrald, J. R., & Roberts, K. (2007). Leading indicators of safety in virtual organizations. *Safety Science*, 45(10), 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*, 50(10), 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*, 45(6), 723-743.
- Hale, A. (2009). Why safety performance indicators? *Safety Science*, 47(4), 479-480. doi:10.1016/j.ssci.2008.07.018
- Hale, A., Guldenmund, F., Loenhout, P. v., & Oh, J. (2010). Evaluating safety management and culture interventions to improve safety: Effective intervention strategies. *Safety Science*, 48, 1026–1035.
- Harms-Ringdahl, L. (2009). Dimensions in safety indicators. *Safety Science*, 47(4), 481-482. doi:10.1016/j.ssci.2008.07.019
- Hinze, J., Thurman, S., & Wehle, A. (2013). Leading indicators of construction safety performance. *Safety Science*, 51(1), 23-28. doi:10.1016/j.ssci.2012.05.016
- Hollnagel, E. (2010). On How (Not) to Learn from Accidents. Retrieved from http://www.uis.no/getfile.php/Konferanser/Presentasjoner/Ulykkesgransking%202010/EH_AccLearn_short.pdf
- Hollnagel, E. (2012). The Health Foundation Inspiring Improvement: Proactive approaches to safety management. [Pamphlet] The Health Foundation. Retrieved from <http://www.health.org.uk/sites/default/files/ProactiveApproachesToSafetyManagement.pdf>.
- Hollnagel, E. (2014). *Safety-I and Safety-II: The Past and Future of Safety Management*. . Ashgate Publishing, Ltd..

- Holt, C. (2014). Safety intelligence & Management Workshop. Presentation at the Safety Intelligence & management Workshop, Dubai, UAE, November 2014
- Hopkins, A., (2012). Disastrous Decisions: The Human and Organisational Causes of the Gulf of Mexico Blowout, CCH Australia Ltd, Australia.
- Hopkins, A. (2009). Thinking about process safety indicators. *Safety Science*, 47(4), 460-465. doi:10.1016/j.ssci.2007.12.006
- Hubbard, D., & Evans, D., (2010). Problems with scoring methods and ordinal scales in risk assessment. *Ibm Journal of Research and Development*, 54, 3, 2:1-2:10.
- IATA (2015). Safety Report 2014. International Air Transport Association. Montreal, Geneva. Retrieved from <http://www.iata.org/publications/Documents/iata-safety-report-2014.pdf>
- 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 (2015). ICAO Safety Report 2015 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. (in press) Critical Review of Safety Performance Metrics, *International Journal of Business Performance Management*
- 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*, 5(2), 104-113, doi: 10.1027/2192-0923/a000083.
- Karanikas, N. (2015b). An introduction of accidents' classification based on their outcome control, *Safety Science*, 72, 182-189. doi:10.1016/j.ssci.2014.09.006.
- Kjellen, U (2000). *Prevention of accidents through experience feedback*. London: Taylor & Francis
- Kruijssen, E. (2013). *Deriving Safety Metrics, from data to intelligence*. Presentation at industry day, The Hague, Netherlands, 19 April 2013.
- Leveson, N. (2011). *Engineering a safer world: Systems thinking applied to safety*. Boston, Mass: MIT Press.
- Maurino, Dan (2005). "[Threat and Error Management \(TEM\)](#)" (PDF). *Coordinator, Flight safety and Human Factors Programme - ICAO. Canadian Aviation Safety Seminar (CASS)*
- 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*, 48, 1520-1529. doi: 10.1016/j.ssci.2010.05.006.
- O'Connor, P., O'Dea, A., Kennedy, Q., & Buttrey, S. E. (2011). Measuring safety climate in aviation: A review and recommendations for the future. *Safety Science*, 49(2), 128-138. doi:10.1016/j.ssci.2010.10.001
- Papadimitriou, E., Yannis, G., Bijleveld, F., & Cardoso, J. L. (2013). Exposure data and risk indicators for safety performance assessment in europe. *Accident Analysis & Prevention*, 60, 371-383. doi:10.1016/j.aap.2013.04.040

- Pasman, H., & 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*, 30, 197-206. doi:10.1016/j.jlp.2013.06.001
- Perrin, E. (2014). Advancing thinking on safety performance indicators. Presentation at the Safety Intelligence & Management Workshop, Dubai, UAE November 2014.
- 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*, 73, 146-166. doi:10.1016/j.ssci.2014.11.018
- Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. *Safety Science* 27 (2–3), 183–213.
- Rasula, J., Vuksic, V. B., & Stemberger, M. I. (2012). The impact of knowledge management on organisational performance. *Economic and Business Review for Central and South-Eastern Europe*, 14(2), 147. Retrieved from Google Scholar.
- Reason, J. (1990). *Human error*. New York: Cambridge University Press.
- Reiman, T., & Pietikäinen, E. (2012). Leading indicators of system safety – monitoring and driving the organizational safety potential. *Safety Science*, 50(10), 1993-2000. doi:10.1016/j.ssci.2011.07.015
- Remawi, H., Bates, P., & Dix, I. (2011). The relationship between the implementation of a safety management system and the attitudes of employees towards unsafe acts in aviation. *Safety Science*, 49(5), 625-632. doi:10.1016/j.ssci.2010.09.014
- Rockwell, T.H. (1959). Safety performance measurement, *Journal of Industrial Engineering*, 10, 12-16.
- Roelen, A.L.C., & Klompstra, M.B., (2012). The challenges in defining aviation safety performance indicators. PSAM 11 & ESREL 2012, 25 - 29 June 2012, Helsinki, Finland.
- Safety Management International Collaboration Group (SM ICG), A Systems Approach to Measuring Safety Performance: A Regulator Perspective (2014), available at: <http://live.transport.gov.mt/admin/uploads/media-library/files/A%20Systems%20Approach%20to%20Measuring%20Safety%20Performance%20-%20the%20regulator%20perspective.pdf>
- Safety Management International Collaboration Group (SMICG) (2013), *Measuring Safety Performance Guidelines for Service Providers*. Retrieved from <http://www.skybrary.aero/bookshelf/books/2395.pdf>
- Salmon, P., McClure, R., & Stanton, N. (2012). Road transport in drift? Applying contemporary systems thinking to road safety. *Safety Science*, 50(9), 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*, 72, 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., & Kerper, S. (2015). Using leading indicators to measure occupational health and safety performance. *Safety Science*, 72, 240-248. doi:10.1016/j.ssci.2014.09.010

Singh, S., Darwish, T. K., Costa, A. C., & Anderson, N. (2012). Measuring HRM and organisational performance: Concepts, issues, and framework. *Management Decision*, 50(4), 651-667. Retrieved from Google Scholar.

Snowden, D. J., & Boone, M. E. (2007). A leader's framework for decision making. *harvard business review*, 85(11), 68.

Stolzer, A. J., Halford, C. D., & Goglia, J. J. (2008). *Safety management systems in aviation*. Aldershot, Hampshire, England ; Burlington, VT: Ashgate. Retrieved from Library of Congress or OCLC Worldcat.

Tinmannsvik, R.K., (2005). Performance indicators of air safety - some results from Swedish aviation. SINTEF, Trondheim, Norway (in Norwegian).

Tump, R. (2014) NLR Flight Test and SMS; what to measure? Presentation at the European Flight Test Safety Workshop 2014, Manching, Germany, 5 November 2014

Underwood, P., & Waterson, P. (2013). *Accident analysis models and methods: guidance for safety professionals*

Wahlström, B., & Rollenhagen, C. (2014). Safety management – A multi-level control problem. *Safety Science*, 69, 3-17. doi:10.1016/j.ssci.2013.06.002

Webb, P. (2009). Process safety performance indicators: A contribution to the debate. *Safety Science*, 47(4), 502-507. doi:10.1016/j.ssci.2008.07.029

Weick, K. E., & Sutcliffe, K. M. (2001). *Managing the unexpected: Assuring high performance in an age of complexity* (1 ed.). San Francisco: JOSSEY-BASS.

Woods, D.D., (2006). Essential characteristics of resilience. In: *Resilience Engineering: Concepts and Precepts*. Ashgate, Aldershot.

Woods, D. D., Branlat, M., Herrera, I., & 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*, 23(2), 97-105. doi: 10.1111/1468-5973.12079

Wreathall, J. (2009). Leading? Lagging? Whatever!. *Safety Science*, 47(4), 493-494. doi:10.1016/j.ssci.2008.07.031

Øien, K., Utne, I. B., & Herrera, I. A. (2011). Building safety indicators: Part 1 – theoretical foundation. *Safety Science*, 49(2), 148-161. doi:10.1016/j.ssci.2010.05.012

Øien, K., Utne, I. B., Tinmannsvik, R. K., & Massaiu, S. (2011). Building safety indicators: Part 2 – application, practices and results. *Safety Science*, 49(2), 162-171. doi:10.1016/j.ssci.2010.05.015

Appendix
Annex 1: Safety process metrics

Areas of measurement	Methods of measurement
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)
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
Ratios of contributions / population (e.g., number of contractors submitted hazard reports / total number of contractors).	
Best practice in safety	Number of activities following best practice

Areas of measurement	Methods of measurement
management over time	(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)
	Number of risks per level (e.g., low risks)
	% of specific risk levels (e.g., % of medium risks)
Culture	Surveys (e.g., extent to which safety culture characteristics are present)
	Analysis of records (e.g., % of workers participating in safety review meetings)
	Comparisons (e.g., changes of safety culture characteristics over time)
Financial	Ratio of budget invested on (specific) safety management activities / quantified costs of events led to losses

Annex 2: Safety outcome metrics

Areas of measurement	Methods of measurement
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