Amsterdam University of Applied Sciences

RAAK PRO Project: Measuring Safety in Aviation

Deliverable: Results from Surveys about Existing Aviation Safety Metrics

November 2016

Authors: Steffen Kaspers, Nektarios Karanikas, Alfred Roelen, Selma Piric, Robbert van Aalst, Robert J. de Boer

Project number: S10931

RAAK PRO Project: Measuring Safety in Aviation

Results from Surveys about Existing Aviation Safety Metrics

Steffen Kaspers[1], Nektarios Karanikas[1], Alfred Roelen[1,2], Selma Piric[1], Robbert van Aalst[1], Robert J. de Boer[1]

[1]Aviation Academy, Amsterdam University of Applied Sciences, the Netherlands
[2]NLR, Amsterdam, the Netherlands

Contents

Kaspers, S., Karanikas, N., Roelen, A.L.C., Piric, S., & de Boer, R. J. (2016). Results from Surveys about Existing Aviation Safety Metrics, RAAK PRO Project: Measuring Safety in Aviation, Project S10931, Aviation Academy, Amsterdam University of Applied Sciences, the Netherlands

1. Introduction

In September 2015, the Aviation Academy of the Amsterdam University of Applied Sciences initiated the research project entitled "Measuring Safety in Aviation – Developing Metrics for Safety Management Systems". The project responds to specific needs of the aviation industry: Small and Medium Enterprises (SME) lack large amounts of safety related data in order to measure and demonstrate their safety performance; large companies might obtain abundance of data, but they need safety metrics of better quality. Therefore, the aim of the project is to identify ways to measure safety in scientifically rigorous, meaningful and practical ways without the benefit of large amounts of data (Aviation Academy, 2014). The research phases are: examination of validity of current safety metrics, exploration of new suitable safety metrics based on existing and alternative models and approaches to safety, generation and validation of a short list of suitable safety metrics, and translation of this knowledge into a web-based dashboard for the industry. The project will last until August 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.

During this first phase of the research (i.e. September 2015 – August 2016) the current views and practices on safety metrics were identified by reviewing state-of-art academic literature, (aviation) industry practice, and documentation published by regulatory and international aviation bodies (Kaspers et al, 2016). This review concluded with the following findings:

1. Safety is widely seen as avoidance of failures and is managed through the typical risk management cycle which includes the stages of hazard identification, risk assessment, risk mitigation and risk monitoring. Under this concept:
    a. Hazards are identified through a spectrum of sources such as mandatory and voluntary reports, internal and external audits, safety investigation reports, and management of change.
    b. Risk assessment is predominately based on probabilistic approaches, which employ estimations of likelihood and severity. Although it is recognised that past performance does not guarantee future performance, likelihoods and severities are estimated with the use of historical data and/or expert judgement, the latter being subject to cognitive biases. In addition, the classification of likelihood and severity classes in risk matrices is not standardised and direct comparisons of risk levels across companies are not feasible.
    c. Risk mitigation or elimination is achieved through barriers of various types (e.g., procedures, technology, training), depending on the available resources and the degree of desired control of the risk.
    d. Risks are actually monitored through the same sources that hazards are identified.
2. Safety metrics can be, conventionally, split in two groups: safety process and outcome metrics.
    a. Safety process metrics are linked with operational, organizational and Safety Management System (SMS) activities. The premise is that better and adequate SMS/safety processes lead to improvement of safety outcomes.
    b. Outcomes are occurrences of any severity category (i.e. accident, serious incident, incident) and they are used by the industry to develop respective indicators (e.g., number of occurrences per aircraft departure) for measuring safety performance. However, the thresholds for incidents and serious incidents are not clearly defined; thus, safety outcomes cannot be directly compared across organizations, and the current taxonomy is differently interpreted. Furthermore, the units of exposure (e.g., departures, miles flown, number of staff) used to develop indicators are not uniform across the industry, and companies choose the ones that confirm their expectations (e.g., correlations between numbers of safety events and operational activity figures). In addition, accidents and incidents are infrequent when

considering the amount of operational activities, therefore they cannot be seen as a useful indication of current safety level.

3. There is a lack of standardization across the aviation industry for the development of safety metrics and there is no explicit reference to quality criteria regarding the design of such metrics. Companies are asked to develop their own safety metrics, a practice that offers flexibility and opportunities for customization. However, this deprives the aviation sector from establishing a common language about safety metrics and perform benchmarks.
4. Safety culture is seen as either an outcome indicator (i.e. a result of safety management) or process indicator (i.e. a reflection and indication of safety management performance). Therefore, there is a lack of consensus whether safety culture needs to be influenced in order to improve safety performance or whether the former is a sort of measurement of the latter.
5. There is limited empirical evidence about the relationship between SMS/safety process and outcome metrics and the link between those often relies on credible reasoning. Such reasoning is principally based on linear safety/accident models, where a cause-effect relation between safety management and safety outcomes is implied. Thus, the relationship between SMS/safety processes and outcome metrics is seen as monotonic in practice and follows a "necessary but not sufficient" logic; a single failure or deviation from a SMS/safety process might not lead to an adverse outcome, but multiple failures (e.g., malfunctioning barriers) or deviations (e.g., incompliance with procedures) are likely to cause unwanted outcomes. Besides the linear accident models, few systemic models have been introduced in literature but they haven't been extensively applied to the industry.
6. Standards have mandated the transition from compliance-based to performance-based evaluations of safety, a concept that is supported by the industry but is not yet backed with specific tools and techniques.

Taking into account the findings from the literature review, this report presents the results from the next part of the research, during which surveys were conducted in order to explore the extent to which the findings from the literature review are reflected in the practice of the partner companies. We examined (1) what, how and why certain safety metrics are used, and (2) whether a monotonic relation between SMS process and safety outcomes metrics is evident; at this stage of the research we did not focus on safety processes at the work floor (i.e. how safety management is actually practiced) and our aim was to evaluate whether SMS processes are linked to safety outcomes.

After formulating the research questions, the report starts with presenting the methodology followed, which included collection and analysis of qualitative and quantitative data. Next, the results of data analysis are presented and followed by a discussion and conclusions. Finally, the report describes the high-level approach for the next steps of the research, which will focus on the development of alternative safety metrics.

2. Research Design

The overarching question that led the design of the research was "To what extent are the results from the literature review evident in industry practice?". In order to answer the main question, sub-questions were formulated and exploratory and causal research were performed through multiple case studies, as elaborated in the following sections of this report. The sub questions (Q1 to Q7) and their correspondence with the literature review findings (section 1 above) are shown in Table 1. It is noted that the research time focused on the metrics used by the companies, thus the findings 1c and 1d were not considered in regarding Q1.

Table 1: Research Sub-questions.

No	Sub-question	Correspondence with literature review findings
Q1	How do the companies perform risk assessments?	1a, 1b
Q2	What types of safety metrics do companies use and are those metrics comparable?	2
Q3	Do the safety metrics used by the companies adhere to the quality criteria mentioned in the literature?	3
Q4	How is safety culture seen in a SMS?	4
Q5	What are the safety paradigms/views used in practice?	5
Q6	Is there a monotonic relationship between SMS process and safety outcomes?	2a, 5
Q7	Are demographic and operational activity figures representative of risk exposure?	2b

Questions Q1 to Q5 were answered through qualitative research, as explained in section 3.2 below; the hypothesis (H1) was that the respective results would confirm the findings from the literature review. In order to answer the questions Q6 and Q7, a causal design was used: the scope was to identify if demographic, operational activity and SMS process data are statistically associated with safety outcomes, whether such associations have a negative or positive direction, and if those are common across the companies surveyed. Based on the aforesaid approach, two main hypotheses were tested in correspondence with questions Q6 and Q7:

H2: There are consistent and similar monotonic relations of SMS process data with safety outcomes across all companies.

In order to judge what type of effect an SMS process has on safety outcomes based on the direction of the relationship, the scope and timeliness of the respective process must be considered. For example, in the cases of safety training and audits, a negative correlation is expected under the argument that more training or audits lead to fewer safety outcomes and vice versa. However, when considering other SMS processes, such as safety reporting and hazard identification, a positive correlation might be expected when the results of the investigation of outcomes retrofit risk assessment; on the other hand, a negative correlation might also reflect that risk assessment does not succeed to increase safety performance, meaning decrease adverse events.

H3: There are consistent and similar monotonic relations (i.e. regardless their positive or negative direction) of demographic and operational activity data with safety outcomes across all companies.

Correlations of operational activity or/and demographic data with safety outcomes (1) over time for each company and (2) across the whole sample when considering respective averages per company, indicate validity of the respective ratios (i.e. monitoring indicators).

3. Methodology

3.1 Sample and Ethics

In order to answer the questions stated in section 2.1 above, the research team interviewed safety managers and professionals from thirteen European aviation companies and also collected numerical data, as explained below in sections 3.2 & 3.3. Companies were represented by one to three safety staff who spoke on behalf of their company. The large companies were represented by safety department

personnel e.g. safety manager, safety specialist and small companies were represented by their safety manager. Out of the 13 companies, seven were large (i.e. >250 employees) and six companies fell under the category of SME (i.e. < 250 people). The participating companies are distributed across four domains: Flights Operators (Flight Ops, N=7), Air Navigation Service Providers (ANSP, N=2), Ground Service Provider (GSP, N=1) and Maintenance, Repair and Overhaul service providers (MRO, N=3). All 13 companies took part in the interviews and ten of those companies provided numerical data.

All data collected during the surveys were treated as strictly confidential and this report includes only anonymised information and data. The company partners will receive individual reports referring to their position in relation to the rest of the sample. Respective Non-Disclosure Agreements were signed for all participating companies.

3.2 Collection and Analysis of Qualitative Data

The interviews were conducted between February and April 2016, according to a predetermined protocol (Appendix 1). The interviews lasted 4 to 6 hours in average; only in one large company the interview duration was limited to 2,5 hours due to time constraints of the company representatives. The interview team consisted of two research team members and one graduate student of the Aviation Academy; one team member was conducted the interview and the other two members were keeping notes. Only in the case of two SMEs which are located outside the Netherlands, the interviews were conducted by one researcher due to travel budget limitations. Eight participant companies allowed the team to record the interviews for future reference and verification of the notes.

Each interview day included four parts:

1. A presentation of the results from the literature review by the research team. This offered ample room for discussing with the safety staff how safety management is practiced and allowed the team and the company representatives to get acquainted.
2. The company explained in more detail how they implement their SMS, giving the opportunity to the research team to ask for clarifications and understand the context of the company before proceeding to the core interview questions.
3. The first part of the interview focused on what, how and why things are measured in regard to safety (see Appendix 1 for the driving questions used in this interview part).
4. The second part of the interview focussed on the SMS elements (ICAO, 2013) that were not explicitly or extensively mentioned by the company during the first interview part (see Appendix 1 for the SMS elements). The scope of this part was to explore what SMS related data companies record but might not directly use in their safety metrics.

The interview notes were cross-checked by all three members of the interview team and when inconsistencies were indicated, the audio files were consulted. For the two companies where only one researcher conducted the interviews, clarifications were provided by email or over the phone. The cross-checks performed was deemed sufficient in order to verify the interview notes; due to time restrictions, the interview notes were not communicated to the interviewees for validation. The verified notes were subject to a template analysis based on the findings of the literature review (Kaspers et al,2016) and the correspondence presented in Table 1.

3.3 Collection and Analysis of Quantitative Data

In order to be able to identify associations of operational activity, demographic and SMS process data with outcomes, we asked the companies to provide data in the form of a data-sheet (Appendix 2).

The creation of the list of data fields was based on the metrics from the literature as those were identified in the previous research stage (Kaspers et al, 2016). The requested data regarded 5 operational activity figures (e.g. departures and miles flown), 12 demographic data fields (e.g. number of staff, number of aircraft), safety outcomes (i.e. safety events in total and number of occurrences, incidents, serious incidents and accidents) and 38 fields covering SMS processes (e.g. hazard identification, SMS documentation updates) from up to 10 years in the past. Specific instructions were not provided to the companies since the fields correspond to data that organisations are familiar with. Clarifications about the requested data were offered to the companies when needed.

Most of the large companies were not able to provide the data requested under the given time frame (i.e. about 1 month). Although SMS process data were available in those companies, they were not always directly linked to safety performance and maintained by the safety department. Therefore, considerable time and resources were needed for the retrieval of the data from several databases. Instead of filling the data sheet, two large companies sent their annual safety dashboards. In these cases, the research team converted the data from the safety dashboards to the respective fields of the datasheet where correspondences were feasible. Also, the data sheets of 3 out of the 10 companies did not include enough data points along time due to their recent business launch and/or relatively recent implementation of a SMS. Consequently, data sets from seven companies were used for statistical tests (Table 2).

Table 2: Sample of quantitative data collection.

	Size		Domain			
	Large (N=7)	Small (N=6)	Flight Ops (N=7)	ATC (N=2)	GS (N=1)	MRO (N=3)
Data-sheets with adequate data points for calculations within the company	2	3	4	1		
Dashboards used for	2		1	1		
Data-sheets with insufficient data points for calculations within the company	1	2	1		1	1

After the collection of datasheets from the companies, raw figures were additionally converted to ratios in order to perform calculations with comparable figures across years for each company (e.g., SMS processes and safety outcomes were divided by activity figures and/or demographic data). The aforementioned conversions resulted in an extensive list of measures (Appendix 3). The researchers tested all available pairs (i.e. Operational Activities – Outcomes, Demographics – Outcomes and SMS processes – Outcomes) as a means to examine all relationships regardless their reference in the literature. Because of the limited sample size, all data were tested with non-parametric correlations. Spearman's coefficient was chosen to explore any monotonic relations of operational activity figures, demographic data and SMS process metrics with safety outcome metrics. It is clarified that the Spearman's coefficient indicates the presence of a monotonic relationship and does not determine the strength of linear associations. The statistical significance level was set to $p=0.05$.

3.4 Results from Qualitative Data Analysis

3.4.1 Risk Assessment and Safety Metrics

Hazard Identification and Safety Metrics

The inputs used by the companies for their risk assessment are shown in Table 3. Those inputs constitute also the basis for measuring safety; the left column of the table refers to the measurements each company uses.

Table 3: Inputs to risk assessment

	Company Size		Activity Domain			
	Large (N=7)	SME (N=6)	Flight Ops (N=7)	ANSP (N=2)	GSP (N=1)	MRO (N=3)
Compliance monitoring	7	6	7	2	1	3
Operational Data [Flight Data Monitoring (FDM) & Air Navigation Service Provider Data Monitoring (ANSPDM)]	5	1	4	2		
Line Operations Safety Audits (LOSA)	2	1	2		1	
SMS Maturity score	2			2		
Feedback from training	1		1			
Voluntary reporting	7	6	7	2	1	3
Safety outcomes [occurrences, (serious) incidents and accidents]	7	6	7	2	1	3
Trends of hazards, events etc. over time	7	6	7	2	1	3

The results in Table 3 show that:

- All companies use compliance monitoring based on the findings from internal and/or external audits, during which it is checked whether the companies follow, standards, legislation, rules, procedures etc. However, one company honestly acknowledged that the value of an audit might be limited "…during an audit everybody puts on their best show, and after the inspectors leave, everybody goes back to normal work".
- Large companies mainly use operational data for their risk assessment. Small service providers do not always have technical capabilities to provide this type of data, and are also not required to collect and analyse this data due to the size of their aircraft (Skybrary, 2016). Flight Data Monitoring (FDM) requires regular downloads of flight data from the aircraft so analysts can retrofit predetermined combination of monitored parameters in a database/computer and observe changes over time across routes, aircraft types etc. Flight data can be downloaded in real time although it is dependable on the available technology of the aircraft and/or air operator. The same concept applies to the Air Navigation Service Providers Data Monitoring (ANSPDM) programs, whereby radar data and radio transmissions are recorded.
- 3 out of the 13 companies use a form of Line Operations Safety Audit (LOSA) as input to their risk assessment. The concept of LOSA is that trained observers evaluate staff during their normal activities. The auditors identify hazards and threats, which might cause negative safety outcomes, they observe the responses of the operators and they provide feedback to the employees and the organization as a means to continuously improve safety. LOSA are internal means of compliance and detection of deviations along with their context, and are different from formal SMS and operational audits conducted by safety assurance staff, authorities, insurance companies etc.
- The two ANSPs assess their SMS regularly with the use of a maturity score, which is a self-scoring method introduced by Eurocontrol (2009).

- One company uses feedback from safety training as input to its risk assessment, where the experiences shared between the instructors and the trainees are used as an information source for the latter SMS process.
- All companies have a system in place where employees can report any safety related case. The interviews indicated that such a formal reporting system in small companies is not consistently used, and coffee table talks among employees comprise a basic source of relevant information. However, for large companies reporting is seen as a valuable resource for their SMS improvement. The use of such a reporting system varies and can be divided in three areas;
    - Identification of hazards.
    - Contextualization of certain situations; for example, when a FDM event is triggered, a voluntary report may be used to add more context to the situation, so the event can be better understood and possible similar event so to be controlled in the future.
    - Indication of safety culture levels; high numbers of voluntary reports are interpreted as an active interest of employees to disclose what is happening at the operational field and an endorsement into the company's just culture.
- All companies interviewed monitor their safety outcomes such as occurrences, (serious) incidents and accidents. However, the participants admitted that the lack of clarity and specific thresholds in the definitions referred in current aviation standards and regulations can result in different interpretations across and within companies.
- All companies look for trends in their data over time, e.g., FDM events, hazards from safety reporting or safety outcomes. The monitoring intervals differ; some small companies look yearly at their numbers and discuss them, while larger companies look at the trends on a monthly basis. However, none of the companies reported the establishment of predetermined alert limits in the monitoring of trends. Hence, trends are evaluated in a qualitative manner; if a trend is recognised, the company might act or not without any reference to predefined limits.

Risk Assessment

After data from the sources mentioned in Table 3 are collected, the risk level is assessed by 11 of the companies with the use of a likelihood-severity matrix. Companies assess the probability and severity based on past cases inside and outside the company or expert judgment when such data is not available or reliable. The resulting risk level determines the urgency and priority amongst risks, which management might reprioritise based on their views or additional contextual information. Finally, unacceptable risks must be mitigated. In addition to this common practice, the information collected during the interviews showed that:

- Nine companies use a 5x5 matrix, whereas the two ANSPs use their own 6x5 design with an additional row/column for undefined/non-assessed risks. Two out of the three MRO companies did not explicitly state the use of such a matrix.
- One air operator stated that the current risk assessment method is completely arbitrary, because the results are highly dependable on the expert who is available each day in order to assess the risk(s).
- One small company felt unsure about the use of its risk matrix due to the lack of data to make probability and severity estimations. The same company mentioned that they are interested in a more objective manner to assess risks and be able to compare those with assessments of other similar companies.

3.4.2 Criteria for safety metrics development

Table 4 presents what criteria companies employ for developing their safety metrics. According to the findings:

- Companies that have established safety metrics follow the guidance of standards (e.g. ICAO Safety Management Manual), own professional knowledge and/or the practices shared in the industry.
- Three large companies try to "measure everything that can be measured" by using all data generated by their systems.
- One small MRO stated that it hasn't established safety metrics, they do not use numerical figures for their risk management and they assess their safety management in qualitative manner.
- One company uses metrics based on a trial and error approach. They look for metrics that are relevant to the process of concern and collect respective data; if the metrics seem suitable, they are maintained and tracked, otherwise they are replaced with new ones. However, criteria for suitability of such metrics were not stated.
- In the same vein, another company acknowledged that they do not have a solid list of safety metrics and the safety metrics change over time.
- Three companies mentioned the SMART criteria (i.e. Specific, Measurable, Agreed/Achievable, Relevant and Time-bound). The company which does not use safety metrics stated that they would use the SMART criteria in the case that they would measure their safety performance; this case has been marked with a "X" in Table 4.

Table 4: Methods for creating safety metrics.

	Company Size		Activity Domain			
	Large (N=7)	SME (N=6)	Flight Ops (N=7)	ANSP (N=2)	GSP (N=1)	MRO (N=3)
Measure what is measurable	3		1		1	1
Based on expert judgement, standards, and professional knowledge	3	4	4	2		1
Trial and error	1			1		
Indicators change over time	1			1		
SMART	2	x		1	1	x

Appendix shows an evaluation of safety metrics of Table 4 against the following criteria found in literature (Kaspers et al, 2016):

- 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.

The evaluation was based on the combination of the information and findings reported in this section and section 3.4.1 above and the results showed that:

- There is no explicit theoretical framework supporting the metrics.
- Most of the metrics are specific and measurable but those characteristics depend on the instrument used for the data collection and the interpretation of the data analysis results.
- Validity of the metrics is only partially met due to factors such as lack of a systemic approach, subjective implementation of the respective tools and ambiguous definitions.
- No metric was completely immune to manipulation
- The practicality and cost-effectiveness of the metrics is dependable on the amount and nature of data collected and analysed in relation with the available resources.
- The reliability of the metrics is not guaranteed due to subjective evaluations most of the metrics require.
- The frequency/periodicity of monitoring is the main factor influencing the sensitivity of metrics to changes of conditions.

3.4.3 Safety culture and models

Nine companies mentioned the importance of culture by referring to one or more types of culture, such as just culture, safety culture or reporting culture (Table 5). However, none of the companies measure their culture consistently; only one ANSP assessed occasionally their safety culture, however the latter not been viewed as a regular safety metric by the specific company.

Table 5: Culture types mentioned by the companies.

	Size		Domain			
	Large (N=7)	SME (N=6)	Flight Ops (N=7)	ANSP (N=2)	GSP (N=1)	MRO (N=3)
Culture (including safety, reporting and just culture)	6	3	5	1	1	2
Safety culture	5	2	3	1	1	2
Reporting culture	1	1	2			
Just Culture	2	1	2	1		

As shown in Table 6, the companies think about safety mainly with a linear, direct cause-effect approach. Only three large companies use both systemic and linear models to analyse incident and accidents, but the choice of the model depends on the resources available; linear models are easier and less costly to apply than systemic ones.

Table 6: Models mentioned by the companies.

	Size		Domain			
	Large (N=7)	SME (N=6)	Flight Ops (N=7)	ANSP (N=2)	GSP (N=1)	MRO (N=3)
Systemic models	3		1	2		
Linear models	6	3	4	2	1	3

3.4.4 Additional information

In addition to the data collected in relation to the research sub-questions, during the interviews the companies expressed their concerns, questions, and ideas about safety metrics as a means to provide the researchers with indicative directions for the next research phase.

Concerns and questions of companies

The companies referred to concerns/questions to be considered in the development of alternative safety measurements and/or techniques, as follows:

- Compliance is not safety.
- How can we interpret statistics in the right manner?
- How do we know whether a SMS process is good?
- How can Safety II be implemented under so many successful movements?
- Occurrence related metrics (e.g., frequency of causal factors) do not reflect the different context of each event.
- If only one barrier remains, does that mean that we are unsafe or that the system worked?

Needs and ideas about safety metrics

The ideas companies stated about the design of alternative safety metrics were about improvement of current practices and test if new safety concepts, as follows:

- Improvement of current practices
    - Living bowtie
    - Data mining
    - Safety culture measurement
    - FDM linked to:
        - Individual flight crew performance
        - Exact location of aircraft
        - Unstable approaches (not recorded when go-around is initiated)
    - Fatigue measurement
    - Cognitive load measurement
- Testing the concepts of:
    - Resilience Analysis Grid
    - Safety II
    - Gap Work as Done vs Work as Imagined

3.5 Results from Quantitative Data Analysis

Table 7 shows the number of pairs (i.e. Operational Activities – Outcomes, Demographics – Outcomes and SMS processes – Outcomes) tested for monotonic relations. The table is divided in three sections corresponding to operational activities, demographics and SMS processes, all of which were tested for correlations with safety outcomes. Within each section, the number of valid pairs are mentioned (i.e. the cases that the data provided allowed statistical calculations) and the significant correlations for those pairs of data [number, (percentage)].

Table 2: Valid pairs tested for monotonic relations

	Operational Activities - Outcomes		Demographics - Outcomes		SMS - Outcomes	
Company	Valid pairs	Significant correlations	Valid pairs	Significant correlations	Valid pairs	Significant correlations
1	4	0, (0%)	0	0, (0%)	25	0, (0%)
2	30	6, (20%)	57	7, (12.3%)	165	19, (11.5%)
3	3	0, (0%)	0	0, (0%)	12	5, (41.7%)
4	36	10, (27.8%)	0	0, (0%)	116	27, (23.3%)
5	232	0, (0%)	188	6, (3.2%)	1292	82, (6.3%)
6	62	8, (12.9%)	48	20, (41.7%)	380	42, (11.1%)
7	72	57, (79.2%)	12	8, (66.7%)	12	8, (66.7%)
Total	439	81 (18.5%)	305	41 (13.4%)	2002	183 (9.1%)

Appendices 5, 6 and 7 report the cases that significant correlations within companies were found. The cells in the corresponding tables include the direction of each correlation (i.e. POS: Positive and NED: negative) and the number of companies for which the data permitted the conduction of valid correlations per case (i.e. sample N). The cells where POS or NEG are followed by a number (i.e. x Number) indicate how many companies had the respective significant correlation; a non-reference to number means that the correlation was found only at one company. The Spearman's coefficient *rho* in the majority of the cases was 1.000 (i.e. positive correlation) or -1.000 (i.e. negative correlation) with a significance of p=0.000; therefore, for space saving reasons the *rho* and *p* values are not reported in the aforesaid Appendices.

In addition to the results within companies, Table 8 shows the significant correlations of the averages of safety outcomes of all severities with activity (i.e. departures and flight hours flown) and demographic data (i.e. number of company staff, full time equivalent of company staff, full time equivalent of contractors, flight hours per pilot, aircraft fleet and aircraft age) across the sample. It is noted that tests for miles flown were not feasible due to limited data. Through those correlations, we aimed at exploring the validity of using demographic or operational activity data as denominators of ratios of adverse safety events, since such ratios are used by the industry in order to compare safety performance.

Table 8: Correlation of averages of activity/demographic data with safety outcomes.

Demographic and Operational Activity Figures (Averages of Companies)	Safety outcomes			
	Serious Incidents	Incidents	Occurrences	All events
Flight Hours	$r_s(6)=0.845$ p = .034		$r_s(5)=0.900$ p = .037	$r_s(6)=0.943$ p = .005
Full Time Equivalent of Contractors		$r_s(4)= -1.000$ p = .000		
Flight Hours per Pilot			$r_s(3)=1.000$ p = .000	$r_s(3)=1.000$ p = .000

The findings presented in Table 8 showed that:

- Increased flight hours' activity is associated with more occurrences, serious incidents and safety events in general.
- The more FTEs are spent by contractors, meaning the more the outsourcing of company activities, the fewer the incidents recorded by the company.
- The more the flight hours' load per pilot the more the occurrences and events in general.

Taking into account that the flight hours was the main variable associated with some types of safety outcomes, we conducted further statistical tests as follows (Table 9):

- Mann – Whitney test was used as a means to explore if the ratios of each event type by flight hours differ between large companies and SMEs. The calculations did not show any statistically significant differences.
- Kolmogorov - Smirnov tests were conducted for the ratios of each event type by flight hours for SMEs; the sample size did not allow the conduction of those tests for large companies and for the categories of serious incidents, occurrences and all events. The results showed significant differences between SMEs regarding their accidents and incidents per flight hours.

Table 9: Differences between and within large companies and SMEs.

Event type / flight hours	Mann – Whitney test between large companies and SME	Kolmogorov – Smirnov tests between SMEs
Accident	p=0.690	p=0.001
Serious Incident	p=0.143	
Incident	p=0.095	p=0.049
Occurrence	p=0.800	
All events combined	p=0.133	

4. Discussion

The results are discussed below in correspondence with the sub-questions of the research and in accordance with the contextual information the researchers collected during the interviews with the company representatives.

4.1 Exploratory Research

4.1.1 How do the companies perform their risk management?

All companies who are obliged to implement a SMS follow the risk cycle included in the SMM (ICAO, 2013) and, consequently, use the risk matrices. However, some companies recognised that the specific risk assessment method is not adequately objective. In the lack of reliable historical data, the estimation of probability and severity of an occurrence is initially performed by a person and, expectedly, is subject to biases, which was acknowledged by few companies. This is also confirmed by literature (Duijm, 2015; Hubbard et al., 2010) and supported by empirical research (e.g., Karanikas & Kaspers, 2016) although guidance to limit the effect of biases exists (e.g., Cooke, and Goossens, 2000). The researchers during the surveys did not collect information about training of experts in companies as a means to deal with effects of cognitive biases in decision making.

SMEs acknowledged a lack of confidence in the risk area limits they have set in their risk matrices since uniformity and standardization is missing in the aviation industry. Therefore, on one hand standards allow companies to tailor their risk matrices based on their operations, but on the other hand little guidance is provided about methods for developing and using such matrices. This potentially leads to a wide variety of methods and measurements, accompanied by their own definitions. This also does not enable a safety risk benchmarking amongst companies; an event for a large company might be just a minor incident when considering the financial implications, but for a SME the same occurrence might be contemplated as of higher severity due to smaller financial yields.

4.1.2 What types of safety metrics do companies use and are those metrics comparable?

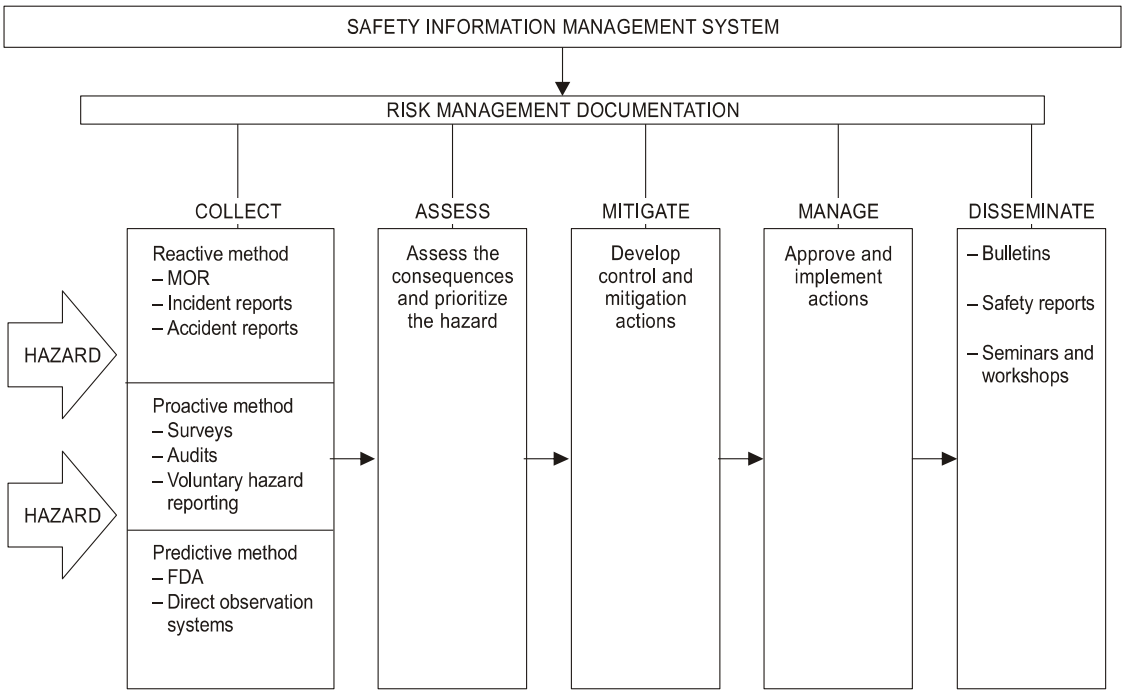
Companies use both SMS/safety process and outcome metrics in the frame of their safety management. Process data are used only to improve safety outcomes without such data being exploited to assess whether individual SMS and safety management processes in general perform adequately. Companies use their safety metrics as sources for identifying hazards that are further subject to risk management under the concept presented in Figure 1 and proposed in the SMM (ICAO, 2013). In other words, the companies' metrics are in the first column "COLLECT" of Figure 1.

All companies collect data about compliance, reporting, outcomes and trends. The results from the survey suggest that:

- Reporting seems to be more formalised at large companies, this possibly attributed to the need to streamline the dataflow. For SMEs, it is easier to share such information since people tend to meet each other more; stories are frequently shared around a coffee table before being reported through formal channels. Regardless of the company size, reporting is highly dependable on perceptions about what is worth to be shared; small, inevitable and normalised deviations might not be reported.
- SMEs have limited access to operational data due to constraints of available aircraft technology and company resources for analysis in combination with the expected volume of data to be processed.
- Large companies look for trends over time in a more systemic manner, at more regular and smaller intervals compared to SMEs. This can be attributed to differences in available resources, volume of operations and staffing levels of safety departments.

- Large companies have generally more data about safety outcomes in terms of raw numbers, but they do not consistently connect and maintain SMS data for use in their safety metrics; hence, it proved cumbersome to identify in their systems the requested data from the research team (e.g., pilot experience might be recorded by the human resources department). SMEs have limited number of safety events, compared to large companies, and they do not also directly associate SMS activities with metrics. However, due to their limited volume of activities in comparison with large companies, it was easier for safety managers and staff at SMEs to fill the datasheet fields requested by the researchers.

Figure 1: Risk Management Process (quoted from ICAO, 2013)



A relation between safety processes and outcomes is expected and assumed, and both safety process and outcome metrics are compared with past figures. Companies seek for improvements when trends over time suggest (e.g., decrease of volumes of voluntary reports, increase of safety events, increase of FDM events of a specific type). Nonetheless, companies have not established any upper and lower control limits about their safety metrics albeit the SMM (ICAO, 2013) requires that companies set goals and alert levels to monitor their safety performance.

Moreover, safety metrics are used both proactively and reactively. Voluntary reports are used in a case-by-case basis for investigating the occurrences reported and derive lessons for the future (a reactive approach). Only one company stated that they use voluntary reporting proactively as a means to identify safety concerns of employees and whether they actively participate in a SMS; this corresponds to a proactive use safety related data. The aforementioned example shows how safety related data might be differently used on the basis of their inherent context or associated numerical figures.

Metrics used by companies do not allow valid comparisons amongst those. First, safety metrics depend on the data collected by each company and are not based on a common standard in terms of data sampling, collection, format, validity and reliability. Even more importantly, as the company representatives mentioned, the widely-used ratios of safety events, and especially the ones of medium and low severity, cannot be directly compared across and within companies due to different interpretations of the respective severity thresholds. For example, a take-off from a taxiway could be classified differently by company analysts depending on their view of the context of the situation: one expert can consider the existence or not of other ground and air traffic at the time of the event and classify the event either as serious incident or just incident respectively; another expert could classify the event as serious incident regardless the conditions, which can be seen as dynamic and not always foreseeable. The context can also affect the points of view of the air operator, the flight crew and the ANSP, all of those possibly classifying the same event differently based on how it had affected their own "process/subsystem". Moreover, each company implements SMS in a different way and develops the respective processes according to their operational profile, needs, resources, size etc. For example, all companies provide safety training to their staff, but the duration, extent and list of topics and the quality of training might vary. Hence, even if a standardised metric of safety training was in place (e.g., percentage of employees successfully completing safety courses, hours spent into safety training per staff annually), it would be difficult to compare the results amongst companies due to the variety of training programs, qualifications of instructors etc.

4.1.3 Do the safety metrics used by the companies adhere to the quality criteria mentioned in the literature?

In general, companies have a rationale behind the development of their safety metrics, but this is not grounded on the whole set of the quality criteria suggested in the literature (Kaspers et al, 2016). Instead, participants follow a pragmatic approach to the indicators used in their SMS and these mainly stem from practice and expert judgment; as soon as metrics seem meaningful to a company, they are maintained and monitored. Amongst the criteria suggested in literature, the "measurable" one was mentioned most often. Even in the case of outcome metrics, the ambiguous definitions across the industry, even within a company depending on the analyst, do not allow a uniformity when classifying events, even within some of the companies it is sometimes hard to reach consensus on classifying a certain event; thus, even the widely-used event rates are not directly comparable among companies, regions etc. The ECCAIRS / ADREP taxonomy (EC, 2014) is an initiative to improve the mandatory reporting by, amongst others, attempting to increase the consistency in the classification of occurrences; however, the use of phrases such as "could have occurred" and "may have been compromised" still offer much space for diverse interpretations.

The "trial and error" approach may indicate that metrics have limited validity. Without predetermined criteria, service providers judge the quality of their current metrics based on expectations and common practice. Interestingly, one SME monitors the frequency of events before deciding to act; this is attributed to the limited number of events that renders statistical calculations invalid, since the sample is highly subject to random noise (e.g., various interpretations, extreme points). Also, the criterion for sensitivity to changes in conditions cannot be ensured with the existing safety outcomes since the latter regard specific findings and events that are not completely repeatable under the dynamic nature of operations.

Few companies mentioned the SMART criteria (Doran, 1981) although those were originally suggested to describe the planning and achievement of management goals, as followed:

- *Specific* – target a specific area for improvement.
- *Measurable* – quantify or at least suggest an indicator of progress.

- *Assignable* – specify who will do it.
- *Realistic* – state what results can realistically be achieved, given available resources.
- *Time-related* – specify when the result(s) can be achieved.

The acronym SMART, used by the companies, is slightly different (i.e. “achievable” replaces “assignable”), but this is a common observation in practice. The SMART criteria do not exactly correspond to the ones suggested in the literature about metrics; validity, cost-effectiveness and the existence of a theoretical framework are not included as part of the criteria to achieve management goals described by Doran (1981). This finding might reflect that companies focus on realising their objectives rather than examining the rigorousness of their metrics.

In general, the results presented in Appendix 4 suggest that no current safety metric fulfils all criteria as identified in the literature (Kaspers, et al, 2016). Few criteria are partially or fully met by current safety metrics (e.g., specific, measurable) and some of those metrics depend on the company resources and measurement instruments. The researchers were not able to trace a specific theoretical framework behind each metric, while it seems that various criteria (e.g., validity, sensitivity to changing conditions, manipulation) were not met in most of the cases. Some explanatory and summative remarks on the results shown in Appendix 4 are as followed:

- Compliance is based on the concept that adherence to the rules ensures a minimum level of safety, but half of the companies stated that safety is more than just compliance. During the discussions, there were connotations that simply following the rules does not guarantee safety. This was interpreted in different ways; first, rules can be realised through various means, the acceptance of the latter being subject to the skills of the auditor. Second, rules do not apply to every situation, since conditions and/or the context of a situation are forevermore changing. Third, there might be situations where rules contradict to each other and final decisions about balancing competing goals rely on the company and/or the end-user.
- Operational data monitoring might be useful to assess frequencies of events but raw data do not capture the context in which these events take place. The context can be provided by reports on the situation identified via the data. However, in the frame of an effective safety management, numbers and coding of events must trigger further exploration of the respective conditions; this depends on available company resources.
- The effectiveness of LOSA depends on the instrument used, the skills of the observer and the perceptions and adaptive behaviour of the subjects.
- The maturity score is a quite abstract and subjective metric. For example, very mature companies might not give themselves the maximum score since they see some room for improvement. Reversibly, companies might overestimate their maturity, since the specific metric is based on self-scoring.
- Reporting that provides context to occurrences is seen as important and can reveal new hazards via the concerns expressed by employees. However, the value of reporting as safety metric is debatable; increased number of reports might indicate that staff trust the company and/or more occurrences happened compared to the past. Also, the quality of the reports determines the opportunities for learning; if only basic information about an occurrence is given, this is just entered in a database and used in statistics. If a report is rich in terms of context, data, views and decisions made, much more may be learned. Furthermore, if companies demand a certain amount of reports from their employees, this might be seen as a requirement for compliance with regulatory requirements that dictate the operation of a ‘voluntary’ reporting process.

4.1.4 How is safety culture seen in a SMS?

Although the companies mention culture as an important element for determining the level of safety, none of the companies measures culture with a predetermined periodicity. The level of safety culture was indirectly indicated through the participation and response of staff to SMS initiatives. For example, safety culture might be indicated through a comparison of FDM triggers with the amount of corresponding voluntary reports. Sometimes safety manager's own perception about the willingness of employees to talk openly indicated a mature safety culture to the companies; although this can provide some indication, it can be subject to biases and more consistent methods and tools should be considered in the assessment of culture. Therefore, companies do not attempt to measure something that they contemplate as a significant part of their safety management. In addition, companies mentioned and linked mostly the reporting and just cultures; other types of cultures [e.g., flexible, informative and learning cultures according to the typology of James Reason (1998)] were not mentioned.

4.1.5 What are the safety paradigms/views used in practice?

The metrics that are used by the companies suggest a focus primarily on negative outcomes, or situations that deviate from normal operations. This would indicate that industry practice is based on traditional views on safety, which is expected since the guidance material from ICAO (2013) refers to linear models such as Reasons' Swiss Cheese (Reason, 1990). However, there is recognition by the companies that the current metrics do not suffice and that compliance alone is not safety. Also, the companies mentioned that they are looking for better metrics to measure safety; some companies look for improved versions of metrics they currently use, and ideas about metrics from other safety paradigms were shared.

Only three companies mentioned the use of systemic models for assessing their safety. The low consideration of newer safety/accident models might be attributed, according to the researchers' knowledge, to the lack of analytical tools that accompany such models or their complexity. At the same time, the companies who stated that they have knowledge about these models, have been yet trying to find practical and manageable indicators that fit the reasoning of the models. Also, companies see some limitations of the newer safety approaches; for example, companies connect Safety II with the measurement of successes, meaning the need to collect much more operational data, thus rendering safety related measurements less practical and costlier compared to traditional metrics. Since concepts such as Safety II have not been yet operationalised through respective techniques, such concerns cannot be judged for their (in)validity.

4.2 Causal Research

4.2.1 Is there a monotonic relationship between SMS process and safety outcomes?

According to the results presented in Appendix 5, the following observations can be made:

1. The significant correlations regard only part of the SMS processes and safety outcomes and a small portion of the sample, and the distribution of associations is highly scattered. No proof was found that all SMS processes had an effect on safety outcomes and the significant associations were found only for few of the participant companies.
2. The results suggest that just the operation of an SMS does not guarantee an effect on safety outcomes; therefore, that other factors, such as the quality of SMS processes, might play an important role. Also, an evaluation of the effectiveness of an SMS against high severity events seems unjustified in the frame of this survey. More specifically:

a. Most of the significant correlations were found for occurrences (i.e. the lowest severity category of safety events) as well as all safety outcomes regardless their severity.
b. Accidents, serious incidents and incidents and their ratios by activity and demographic figures were associated with a very few SMS processes.
c. Since all companies reported many more low severity outcomes than events with high impact, it can be claimed that the correlations regarding all safety events reflect actually the occurrences.
d. The aforementioned picture implies that only some SMS processes at few companies had a visible effect on low severity events which are more frequent and reflect safety performance at shorter intervals.

3. There were 33 negative and 124 positive correlations between SMS process and safety outcomes. However, in 59 cases of all correlations the data regard a single company which was the only one that provided adequate data, so the results cannot be deemed as representative of the whole sample. Nevertheless:
    a. The negative correlations sporadically regarded numbers or ratios related to staffing of the safety department, internal audits, safety training, safety surveys and hazard identification. Although due to the limited sample those associations do not reflect the situation at all companies surveyed, the aforementioned areas of SMS processes were influential on safety outcomes of low severity mostly for a single company. It is noticed that a negative correlation between SMS processes and safety outcomes can be considered as a positive case only when outcomes decrease over time (i.e. increased SMS activity leads to fewer safety events); in case that, under a negative correlation, events increase over time, the SMS can be contemplated as insufficient (i.e. fewer SMS processes lead to more adverse outcomes), if not a contributing factor to decreased safety performance.
    b. Most of the positive correlations were found for the safety reporting and risk assessment processes, the interpretation of those associations being dependable on the timeliness of those processes. The aforementioned SMS activities are performed continuously, so a distinction between a “positive reactiveness” (e.g., more risk assessments occur due to more outcomes) and “negative proactiveness” (e.g., more risk assessments lead to an increase of adverse events) is not directly evident. As discussed in sections 4.1.2 and 4.1.3 above, contextual information is of paramount importance in order to interpret such results correctly. The latter was not feasible during this part of the research due to time limitations, but it will be considered at the next research phases.

The arguments No 2 to No 5 presented above, taking into account the overall picture as stated in observation No 1 above, suggest that hypotheses H2 (i.e.” There are consistent and similar monotonic relations of SMS process data with safety outcomes across all companies”) is partially rejected due to the limitations imposed by the sample size. In addition to the latter factor, the researchers contemplate that the diverse ways that SMS processes are implemented across the industry and over time and the different interpretations of outcome thresholds, as discussed in section 4.1.2 above, affected the results and did not allow completely valid comparisons within and between companies.

4.2.2 Are demographic and operational activity figures representative of risk exposure?

Correlations between operational activities and safety outcomes

The results presented in Appendix 6 do not suggest a consistent picture within companies. Some activity data related to departures, miles flown and flight hours were associated with all safety events, incidents and serious incidents, but in the majority of the cases those findings regarded only one company out of the whole sample. Only in seven cases the associations of flight hours related data with some types of safety outcomes were found for two companies. Interestingly, accidents were not represented in the

significant correlations with operational activities, although annual reports published by regional and international bodies use rates of accidents as a means to depict safety performance (e.g., EASA, 2016); perhaps, the large sample that such reports include might render the use of accident ratios meaningful, but the results of our survey showed that those ratios might not be representative of safety performance at the company level or, in general, when analysing small samples. The latter is also supported by the fact that we did not observe any association between operational activity data and number of accidents when considering averages across the sample (Table 8 in section 3.5).

Furthermore, in the case of flight hours, the correlations with outcomes were found interchangeably positive or negative depending on the denominator and the company, whereas in few cases the same correlation was found negative for one company and positive for another. This observation might reflect also the dissimilarities in the interpretation of safety outcome definitions, as discussed in sections 4.1.2 and 4.2.1 above, or/and the differences regarding the effectiveness of safety management in those companies; a positive correlation between activity and outcome data indicates that safety management is not improving (i.e. as safety management activities increase, safety outcomes increase too and vice versa), whereas a negative correlation signals that safety management either performs either as expected (i.e. when outcomes decrease over time) or poorly (i.e. when outcomes increase over time).

As shown in Table 8 (section 3.5), monotonic relations were found across the companies regarding flight hours and flight hours per pilot with safety outcomes, the accidents excluded, thus suggesting that the specific type of operational activity might be a more valid exposure measurement than departures and miles flown. By nature, departures do not reflect the total load imposed to company staff (e.g., time pilots fly or maintenance requirements based on the hours that aircraft operate), and miles flown are not also directly related to the total load due to a variety of factors such as aircraft capabilities, flight plans and fuel efficiency policies (e.g., the same distance might be covered in shorter or longer time based on the air traffic and average flying speed). The findings of our study are aligned with the approach of Karanikas (2015b), who showed a relation of task load expressed in total flight hours per employee with rates of events attributed to human error.

Correlations between demographics and safety outcomes

The picture in Appendix 7 is even more distorted in comparison with the findings included in Appendix 6 regarding the relationship between operational activity figures and safety outcomes. The correlations found were highly dependable on the denominators used in the safety outcomes; for example, the average aircraft age was positively correlated with number of occurrences and the ratio of occurrences by flight hours, but negatively correlated with the ratios of occurrences by miles flown and departures. Hence, under the expectation that the higher the age of the aircraft the more the occurrences, it seems that flight hours can act as more representative denominators, whereas miles and departures may be confounded by type of operations; this is aligned with the claim made in the previous paragraphs of this section.

Furthermore, the number of company employees was positively correlated with occurrence data, but negatively associated with incidents and all safety events regardless severity. Although those differences do not refer to the same company, they suggest that the use of raw demographic data alone cannot render respective indicators valid. In conjunction with the discussion of the results of Appendix 6 and the paragraph above, ratios of activity figures, and especially flight hours, by demographic data can be more valid representations of risk exposure in comparison with net numbers of operational activities or demographics.

Taking into account the overall picture as discussed above, the researchers claim that the hypotheses H3 (i.e." There are consistent and similar monotonic relations of demographic and operational activity data with safety outcomes across all companies.") is partially rejected since, as stated in section 4.2.1 above, the limited sample size, and the different interpretations of outcome thresholds might have affected the results and did not allow completely valid comparisons within each company.

4.2.3 Overall evaluation of causal research results

From the numerical analysis of the data sample, consistent correlations between operational activity figures, demographic data, SMS process data and safety outcomes could not be established. The correlations in the sample have a wide variety, and there were no correlations supported by all usable datasets. Only part of the datasets resulted to significant correlations for specific combinations of data, and in some cases, there were both positive and negative correlations for the same pair of variables in the sample. Since hypotheses H2 and H3 cannot be fully confirmed, the current practices in safety performance measurement seem of limited validity. The partial rejection of hypotheses H2 and H3 is aligned, and indirectly validated, with the concerns of the companies about the existing safety metrics and their needs for better / alternative ones. Nevertheless, the diverse and, occasionally, contradictory findings from the quantitative analysis might be attributed to:

- The different interpretations of thresholds of safety outcomes.
- The implementation of SMS processes in various ways, due to which the data points reflected different contexts of the companies and changes over time, this probably distorting the results.
- The limited value of the linear approach to safety, as suggested by the models widely used by the industry.

5. Conclusions

The results of the analysis of qualitative data partially verified the findings from the literature review performed before the surveys (Kaspers et al, 2016). On one hand, it was confirmed that:

- Safety is managed through the risk management cycle described in standards, and companies acknowledge the limitations of current risk assessment techniques.
- The safety data collected by the companies retrofit the risk assessment and safety assurance processes.
- Safety outcomes are used as measurement of safety performance, but the definitions of their severities are ambiguous.
- Accidents and incidents are infrequent events, especially at small companies, and cannot constitute reliable measurements of safety performance.
- Companies do not use predefined quality criteria for the design of their safety metrics; each company uses metrics that are specifically tailored to their organisation in terms of type of operations and availability of data.
- Traditional approaches are used for safety management, and most of the companies follow linear models such as the Swiss cheese model and bowties. Few companies already explore newer methods and approaches to safety based on systemic models.
- Companies recognise that better indicators are necessary in the future, and there are also concerns about the feasibility of establishing metrics of high quality in the future.
- Safety culture is seen as important part of safety management.

On the other hand, the research, contradictory to the expectations raised from the literature review, revealed that:

- Current safety metrics are not grounded on sound theoretical frameworks and, in general, do not fulfil the quality criteria as proposed in literature.
- Safety culture is not a consistent part of safety metrics and, therefore, not assessed.
- The companies collect data related to their SMS processes, but such data are not associated with SMS metrics; hence, some of the processes are performed but not measured.
- The data used differ across companies depending on own perceptions, safety models used implicitly or explicitly, and available resources.
- SMS assessment is yet based on a compliance-based approach, whereas standards require the transition to a performance-based evaluation.
- Few, diverse and occasionally contradictory correlations were found between SMS process and outcome metrics. This picture might be attributed to a combination of factors, which are linked to the limitations of a linear approach and the different ways SMS processes are implemented and safety outcomes are classified.

Especially regarding the results from the causal research, those led to the partial rejection of hypotheses H2 and H3 regarding the consistency and similarity of monotonic relationships of SMS process, operational activities and demographics with safety outcomes. Due to the limited sample size (i.e. number of participating companies and data points per company) we do not claim external validity of the results and we could not fully reject those hypotheses.

In overall, the findings of this study indicate the need to move towards the development of metrics that will be more representative of SMS processes and safety outcomes and will allow valid comparisons over time and across the industry. Based on the results of this research phase, the justification of the current project does not only stem from a need to improve scientific knowledge on the topic of aviation safety metrics, but it is also jointly supported by the concerns and needs of the industry and the findings of the analysis of numerical data collected in this research phase.

6. Next Steps

Based on the above, at the next phase of the research the goal is to determine alternative safety metrics which are suitable for SME's and fulfil the quality criteria (Kaspers et al, 2016). The team will research and apply the following:

- SMS and safety processes representation/modelling with systemic models (e.g. Leveson 2011, Hollnagel 2012). Such models have been reported to been reportedly successfully applied to safety assessments across a large array of industries (Leveson et al., in press; Macchi et al., 2009), but they have not yet been operationalised in the area of safety performance, although suggestions in that direction have been made (Leveson 2015).
- Use of the representations/models to depict various "system/process states" according to standards (i.e. ideal system), company policy (i.e. documented system), management practice (i.e. implemented system) and end-user practice (i.e. operationalized system).
- Measurement of the distances between the various system/process states and their effects on safety outcomes by examining different classification taxonomies of the latter (e.g., Karanikas, 2015a) and under the concept that a large gap between work-as-imagined and work-as-done indicates a drift into failure and leads to decreased performance (Hollnagel, 2014; Dekker, 2011).
- Measurement of common high-level factors that can explain the aforementioned distances/effects, such as the degree of coupling between processes (Perrow, 1984), validity of design assumptions (Leveson, 2015), safety culture development (Karanikas et al, 2016), efficiency thoroughness trade-off (Hollnagel, 2009), unruly technology, scarcity of resources and competing goals (e.g., Rasmussen, 1998; Dekker, 2011), and views on human errors (Dekker, 2015).

Acknowledgments

The research team would like to expresses their deep thanks to:

- The companies which participated in the surveys (in alphabetical ascending order of company): Helicentre, JetSupport, KLM (Royal Dutch Airlines), KLM Cityhopper, Life Line Aviation, LNVL (Air Traffic Control the Netherlands), MUAC (Maastricht Upper Area Control Centre), Olympus Airways, SAMCO, Sky Service Netherlands, Transavia.
- The Aviation Academy graduates, who participated in this phase of the research in the frame of their graduation projects: Hayo Brugmans, Marijn van der Deure, Mitchel Nauman and Shivaid Ishaak.
- The 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): CAA NZ: Charlotte Mills, EASA: John Franklin, Kindunos: John Stoop, KLM Cityhopper: Ewout Hiltermann, Klu / MLA: Ruud Van Maurik, Purdue University: Julius Keller, Team HF: Gesine Hofinger, TU Delft:Alexei Sharpanskykh.

References

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

Cooke, R.M., Goossens, L.H.J. (2000). Procedures guide for structured expert judgement, EURATOM document EUR 18820EN, European Communities.

Doran, G. T. (1981). "There's a S.M.A.R.T. way to write management's goals and objectives". Management Review. AMA FORUM. 70 (11): 35–36.

Duijm, N.J., (2015). Recommendations on the use and design of risk matrices. Safety Science, 76, 21-31

EASA. (2016). Annual Safety Review. Cologne: European Aviation Safety Agency.

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

Eurocontrol. (2009) ATM safety framework maturity survey for ANSPs, 3rd SAFREP TF report to Provisional Counsel Appendix 1 2009.

Dekker, S. W. A., (2015) Safety Differently: Human Factors for a New Era. New York: CRC Press.

Dekker, S. W. A. (2011) Drift into Failure: From Hunting Broken Components to Understanding Complex Systems. Farnham, UK: Ashgate Publishing.

Skybrary (2016) Flight Data Monitoring. Retrieved from http://www.skybrary.aero/index.php/Flight_Data_Monitoring

Hollnagel, E., (2009). The ETTO Principle: Efficiency-Thoroughness Trade Off. Farnham: Ashgate.

Hollnagel, E. (2012). FRAM: The Functional Resonance Analysis Method. Modelling Complex Socio- technical Systems. Ashgate: Farnham Surrey UK

Hollnagel, E. (2014). Safety-I and Safety-II: The Past and Future of Safety Management Ashgate: Farnham Surrey UK

Hollnagel, E. (2015) RAG – Resilience Analysis Grid. Retrieved from http://erikhollnagel.com/onewebmedia/RAG%20Outline%20V2.pdf

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.

ICAO (2013). Doc 9859, Safety Management Manual (SMM) (3rd Ed.) International Civil Aviation Organization. Montréal, Canada.

Karanikas, N. (2015a), An Introduction of Accidents' Classification Based on their Outcome Control, Safety Science, 72, pp. 182-189.

Karanikas, N. (2015b). 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), pp. 104-113.

Karanikas, N & Kaspers, S. (2016). Do Experts Agree When Assessing Risks? An Empirical Study. Proceedings of the 50th ESReDA Seminar, 18-19 May 2016, Seville, Spain.

Karanikas, N., Soltani, P., de Boer R. J. & Roelen A.L.C. (2016). Safety Culture Development: The Gap Between Industry Guidelines and Literature, and the Differences Amongst Industry Sectors, in Arezes, P. (ed.), Advances in Safety Management and Human Factors, Proceedings of the AHFE 2016 International Conference on Safety Management and Human Factors, July 27-31, 2016, Walt Disney World®, Florida, USA, Springer

Kaspers, S., Karanikas, N., Roelen, A.L.C., Piric, S., & de Boer, R. J. (2016). Review of Existing Aviation Safety Metrics, RAAK PRO Project: Measuring Safety in Aviation, Project S10931, Aviation Academy, Amsterdam University of Applied Sciences, the Netherlands.

Leveson, N. (2015). A systems approach to risk management through leading safety indicators. Reliab Eng Syst Saf, 136, pp. 17–34

Leveson, N. (2011). Engineering a safer world: Systems thinking applied to safety. Boston, Mass: MIT Press.

Leveson, N., Samost, A., Dekker, S.W.A., Finkelstein S., and Raman, J. (in press) A Systems Approach to Analyzing and Preventing Hospital Adverse Events. Retrieved from http://sunnyday.mit.edu/papers/CAST-JPS.pdf

Macchi, L., Hollnagel, E., Leonhard, J. (2009) Resilience Engineering approach to safety assessment: an application of FRAM for the MSAW system. EUROCONTROL Safety R&D Seminar, Oct 2009, Munich, France. EUROCONTROL, 12 p., 2009. <hal-00572933>

Perrow, C. (1984). Normal Accidents: Living with High-Risk Technologies. Basic Books : USA

Rasmussen, J. (1998). Risk management in a dynamic society: a modelling problem. http://dx.doi.org/10.1016/S0925-7535(97)00052-0

Reason, J. (1998). *Achieving a safe culture: theory and practice*. Work Stress 12(3), 293–306 (1998)

Reason, J. (1990). *Human error*. New York: Cambridge University Press.

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.

Appendix 1: Survey

Outline of surveys at company partners

RAAK PRO research project: Aviation Safety Metrics Phase 1: Defining short valid list of current safety metrics Activity: Surveys at company partners (Feb – Apr 2016)

Outline of the surveys at company partners Day 1:

- Welcome, mutual introduction between researchers and company representatives (10 minutes)
- Presentation by the research team on main topics of the 1st project report (Review of Existing Aviation Safety Metrics) and the progress of the project (20 – 30 minutes including questions)
- (Optional): Presentation of the SMS of the company (20 – 30 minutes including questions and answers)
- Coffee brake (15 – 20 minutes)
- Interview No 1 with safety manager and/or other appointed safety staff. (up to 1.5 hours).

Indicative questions:

- How is safety measured in the company (specific methods, list of metrics) o What is the perceived value of those methods and metrics? What do they represent?
- How are results of safety metrics are used?
- How is safety performance demonstrated to authorities?

Difference analysis: The research team will compare the metrics the company uses with the ones referred in the literature (Lunchbreak – 1 hour)

Interview No 2 with safety manager and/or other appointed safety staff: Discussion about the results of the differences. What might be the reasons of the differences? (up to 1.5 hours). Discuss strategy on collecting raw data samples.

Day 2 (Days 2 & 3 for large companies)

Researchers collect representative samples of raw data on metrics the company uses, including frequency, resources, means, methods, and forms.

Researchers collect representative samples of raw data that can be exploited in safety metrics included in the first project report but the company does not currently use.

Day 1 main/driving questions

1. How do you demonstrate to the authorities and the public that you are/safe(ly)?
2. How do you measure your safety performance?
3. How do you use your safety performance results? (are they connected to the safety objectives and/or SMS improvements?) (threshold limits? per time period)

4. Are there any other indicators that you know, but do not use? (Why?)
5. How have you chosen your safety performance indicators (criteria)?
6. Do you collect any data about other SMS activities that that you do not use in your safety performance measurement?

(Gap analysis – Lunch break)

7. Do you collect data about ….? Do you evaluate …? Do you have metrics about ….? Why? Do you know …? Could you link … with safety performance? Does it make sense? What do you think of …? Etc.

Safety pillar / process	**Letter**	**Checklist**
Management commitment and responsibility	A	
Safety accountability of managers	B	
Appointment of key safety personnel (staffing)	C	
Emergency response	D	
SMS Documentation	E	
Hazard Identification	F	
Risk assessment	G	
The management of change	H	
Continuous improvement of the safety system	I	
Training and education	J	
Safety communication	K	
Safety assurance	L	

Measurement types		
Raw numbers/ counting	1	
Percentages	2	
Ratios	3	
Time Frequencies	4	
Average Values	5	
Other	6	
	7	

Appendix 2: Data Sheet

		Annual Average									
Data	Average	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006
General Information											
Departures											
Miles Flown											
Flight Hours											
Number of Company Staff											
Full Time Equivalent (Company)											
Full Time Equivalent (Contractors)											
Experience of Flight Crews (Flight Hours)											
Hours Flown / Pilot											
Experience of Ground Staff (Years)											
Ratio of Company Staff Turnover											
Ground movements											
Km Driven											
Aircraft Fleet											
Aircraft Age (Years)											
Ratio of Aircraft Fleet Unavailability (e.g., scheduled & unscheduled maintenance)											
Major Ground Equipment Age (Years)											
Ratio of Major Ground Equipment Unavailability (e.g., scheduled & unscheduled maintenance)											
Safety Events											
Number of All Safety Related Events											
Number of Occurrences											
Number of Incidents											
Number of Serious Incidents											
Number of Accidents											
Safety Staff											
Number of Safety Staff											
Full Time Equivalent Safety Staff Spends on SMS											
Number of Safety Staff Changed											
Improvements											
SMS updates											
SOPs, procedures, rules etc. updates											
Number of External Audits											
Findings from External Audits											
Number of Internal Audits											
Findings from Internal Audits											
Number of Internal Safety Reviews / Meetings											
Days for Implementing Decisions from Internal Safety Reviews / Meetings											
Number of Safety Meetings with External Organizations											
Number of Safety Conferences, Workshops etc. Attended											
Number of Safety Surveys											
Ratio of Targeted Population Participated in Safety Surveys											
Number of Safety Studies Accomplished (in addition to Safety Surveys)											
Safety Training & Education											
Number of Safety Training Sessions Completed											
Hours per Safety Training Session											
Ratio of Staff Attending Safety Training											
Ratio of Staff Passing Safety Training Exams on 1st Attempt											

Safety Communication											
Number of Safety Bulletins, Notices etc.											
Times of Safety Communication (each communication might include 1 or more safety messages, posters etc.)											
Hazard Identification											
Number of Safety Reports Submitted by Company and Contractor Staff (e.g., Air Safety Reports, Hazard Reports)											
Number of Safety Reports Followed-Up / Feedback Provided											
Number of Hazards Identified from Sources Except Safety Reports (e.g., Safety Investigations, Safety Audits, Safety Observations)											
Safety Risk Assessment & Mitigation											
Number of Total Risk Assessments Performed											
Number of Risk Assessments Initially Rated as Low											
Number of Risk Assessments Initially Rated as Medium											
Number of Risk Assessments Initially Rated as High											
Number of Risk Assessments Initially Rated as Unacceptable											
Number of Low Risks in the Registry (after assessment & mitigation)											
Number of Medium Risks in the Registry (after assessment & mitigation)											
Number of High Risks in the Registry (after assessment & mitigation)											
Days Between Hazard Identification and Risk Assessment											
Days Between Risk Assessment & Implementation of Risk Mitigation Measures											
Emergency Response											
Number of Emergency Response Exercises											
Hours Spent on Each Emergency Response Exercise											
Number of Emergency Response Planning Updates											

Appendix 3: Extended data-sheet

Nominators (numerator)	Denominator					
Miles Flown	dep per staff	dep per FTE	dep per AC			
Full Time Equivalent (Contractors)						
Ratio of Company Staff Turnover						
Ratio of Aircraft Fleet Unavailability (e.g., scheduled & unscheduled maintenance)						
Major Ground Equipment Age (Years)						
Ratio of Major Ground Equipment Unavailability (e.g., scheduled & unscheduled maintenance)						
Number of All Safety Related Events	per dep	per miles	per fh	per staff (FTE)		
Number of Occurences	per dep	per miles	per fh	per staff (FTE)		
Number of incidents	per dep	per miles	per fh	per staff (FTE)		
Number of Serious incidents	per dep	per miles	per fh	per staff (FTE)		
Number of Accents	per dep	per miles	per fh	per staff (FTE)		
Number of Safety Staff	per dep	per miles	per fh	per staff (FTE)		
Full Time Equivalent Safety Staff Spends on SMS	per dep	per miles	per fh	per staff (FTE)		
Number of Safety Staff Changed	per number of safety staff					
SMS updates						
SOPs, procedures, rules etc. updates						
Number of External Audits						
Findings from External Audits	per audit					
Number of Internal Audits	per safety staff					
Findings from Internal Audits	per audit					
Number of Internal Safety Reviews / Meetings	per safety staff					
Days for Implementing Decisions from Internal Safety Reviews / Meetings						
Number of Safety Meetings with External Organizations	per safety staff					
Number of Safety Conferences, Workshops etc. Attended	per safety staff					
Number of Safety Surveys						
Ratio of Targeted Population Participated in Safety Surveys						
Number of Safety Studies Accomplished (in addition to Safety Surveys)						
Number of Safety Training Sessions Completed						
Hours per Safety Training Session						
Ratio of Staff Attending Safety Training						
Ratio of Staff Passing Safety Training Exams on 1st Attempt						
Number of Safety Bulletins, Notices etc.	per safety staff	per total staff				
Times of Safety Communication (each communication might include 1 or more safety messages, posters etc.)	per safety staff	per total staff				
Number of Safety Reports Submitted by Company and Contractor Staff (e.g., Air Safety Reports, Hazard Reports)	per dep	per miles	per fh	per safety staff	per total staff	
Number of Safety Reports Followed-Up / Feedback Proved	per report submitted	per dep	per miles	per fh	per safety staff	per total staff
Number of Hazards entified from Sources Except Safety Reports (e.g., Safety Investigations, Safety Audits, Safety Observations)	per report submitted	per dep	per miles	per fh	per safety staff	per total staff
Number of Total Risk Assessments Performed	per safety report submitted	per safety staff				
Number of Risk Assessments Initially Rated as Low	as ratio					
Number of Risk Assessments Initially Rated as Medium	as ratio					
Number of Risk Assessments Initially Rated as High	as ratio					
Number of Risk Assessments Initially Rated as Unacceptable	as ratio					
Number of Low Risks in the Registry (after assessment & mitigation)	as ratio					
Number of Medium Risks in the Registry (after assessment & mitigation)	as ratio					
Number of High Risks in the Registry (after assessment & mitigation)	as ratio					
Days Between Hazard entification and Risk Assessment						
Days Between Risk Assessment & Implementation of Risk Mitigation Measures						
Number of Emergency Response Exercises						
Hours Spent on Each Emergency Response Exercise						
Number of Emergency Response Planning Updates						

Appendix 4: Safety metrics used against quality criteria

Criterion	Company metrics							
	Compliance Monitoring	Operational Data Monitoring	LOSA	Maturity score	Feedback from training	Voluntary reporting	Safety outcomes	Trends
Based on a thorough theoretical framework	Indicative combination of theories: Normative behaviour, Taylorism, Laws of Cause-Effect, Quality management							
Specific in what is measured	Yes, but dependable on the instrument used.	Yes, based on predefined parameters.	Yes, but dependable on the instrument used.	Yes, based on the level of improvement efforts.	No.	Yes. (i.e. volume of reports, predetermined coding of factors).	Partially. Ambiguous thresholds.	Yes, mainly changes over time.
Measurable, so to permit statistical calculations	Yes. Differences over time and across various departments (i.e. assuming that the same auditing tool will be used).	Yes (i.e. assuming that the combination of monitored parameters will not change).	Yes. Differences over time and amongst individuals, groups etc.	Yes.	No.	Yes (e.g. number of reports, frequencies of contributing factors).	Yes.	Yes.
Valid (i.e. meaningful representation of what is measured)	Partially. Assessment of individual auditing topics do not address interactions, interdependences and effects of competing goals among various business functions.	Partially. Selection of parameters based on experience and possibly separated from context.	Partially. Dependable on variable individual performance, team dynamics and operational conditions.	Partially. Aggregation of maturity scores of individual functions in an overall score is questionable.	(Not applicable)	Partially. Dependable on the context.	Partially due to ambiguous thresholds.	(Not applicable – dependable of what is monitored).
Immune to manipulation	Partially. Audits are preannounced and organizations get prepared to demonstrate compliance; possibly not capturing daily levels of compliance.	No	Partially. Subjects might adapt their practices in the presence of the observer.	Partially. Organizations might accelerate resolution and documentation of pending issues just before the assessment.	Partially if feedback is provided only in unstructured ways (i.e. without use of evaluation forms).	No	No	Partially. Alert limits not always set or might be changed to accommodate inconvenient trends.

Criterion	Company metrics							
	Compliance Monitoring	Operational Data Monitoring	LOSA	Maturity score	Feedback from training	Voluntary reporting	Safety outcomes	Trends
Manageable – practical (i.e. comprehension of metrics by the ones who will use them)	Partially. Internal auditors are usually aware of special conditions of the company. External auditors might adhere only to the check list topics.	Dependable on skills of the analyst.	Dependable on the skills of the observer and the clarity of the instrument.	Dependable on the instrument.	Yes, in the case of unstructured discussions. Dependable on the clarity of the evaluation form if present.	Dependable on the volume of the reports and the number/nature of the coding fields in combination with available resources for analysis.	Partially, dependable on the level of agreement in the classification.	(Not applicable – dependable of what is measured).
Reliable, so to ensure minimum variability of measurements under similar conditions	No, dependable on the auditor.	Yes.	No, dependable on the observer. (although can be improved by training and interrater-consent)	No, dependable on the assessor. (although can be improved by training and interrater-consent)	No, subject to interpretations.	No, subject to interpretations.	Partially due to ambiguous definitions and possible different interpretations over time and across individuals.	(Not applicable – dependable of what is measured).
Sensitive to changes in conditions	Dependable on duration and periodicity.	Yes.	Dependable on the duration and periodicity.	Dependable on periodicity.	No.	No, each report regards a specific set of conditions.	No, since only the actual severity is measured.	Dependable on the volume of data and frequency of their collection.
Cost-effective, by considering the required resources	Dependable on the extent, depth and frequency of checks.	Dependable on available technology and company resources.	Dependable on the extent, depth and frequency of checks.	Dependable on the extent, depth and frequency of checks.	Yes	Dependable on available technology and company resources.	Yes	Dependable on available technology and company resources.

Appendix 5: Significant Correlations between SMS and Outcome Data

	Safety Staff per staff FTE	Number of Safety Staff Changed	FTE Safety Staff Spends on SMS	Safety Staff Spends on SMS per dep	FTE_Safety_Staff_Spends_on_SMS_per_fh	FTE Safety Staff Spends on SMS per staff FTE	SMS updates	SOPs procedures rules etc. updates	Number of External Audits	Findings from External Audits	Findings per external audit
Number of All Safety Related Events		Pos 3			Neg 3						
All Safety Related Events per dep											
All Safety Related Events per miles											
All Safety Related Events per fh											
All Safety Related Events per staff FTE											
Number of Occurences			Pos 2	Neg 1	Pos & Neg 3	Pos 1	Pos 3	Pos 2		Pos 4	Pos 3
Occurences per dep	Pos 2				Neg 2				Pos 1		
Occurences per miles	Pos 1								Pos 2		
Occurences per fh											
Occurences per staff FTE											
Number of Incidents				Neg 1	Neg 2						
Incidents per dep					Neg 1						
Incidents per miles											
Incidents per fh											
Incidents per staff FTE											

	Number of Internal Audits	Findings from Internal Audits	Findings per internal audit	Number of Internal Safety Reviews Meetings	Number of Safety Training Sessions Completed	Ratio of Staff Attending Safety Training	Hours_per_Safety_Training_Session	Number of Safety Bulletins Notices etc	Times of Safety Communication each communication might include	Safety_Communication__per_total_staff
Number of All Safety Related Events	Pos x2 3	Pos 4	Pos & Neg 3			Neg 2				Pos 4
All Safety Related Events per dep	Pos 3									Pos 3
All Safety Related Events per miles			 1							Pos 2
All Safety Related Events per fh	Pos 2		Neg 3							Pos 4
All Safety Related Events per staff FTE			 3		Neg 2		Neg 2			
Number of Occurences	Pos 2		Neg 3	Pos 3				Pos 3	Pos 3	Pos 3
Occurences per dep						Neg 1				
Occurences per miles						Neg 1				
Occurences per fh	Pos 2		Neg 3							
Occurences per staff FTE										
Number of Incidents	Pos 2									Pos 3
Incidents per dep	Pos 2									Pos 2
Incidents per miles										Pos 2
Incidents per fh										Pos 3
Incidents per staff FTE										
Number of Serious Incidents										
Serious Incidents per dep	Pos 1									
Serious Incidents per miles										
Serious Incidents per fh										

	Number of Safety Bulletins Notice per total staff	Number_of_Safety_Surveys	Number of Safety Reports Submitted by Company and Contractor Sta	Safety_Reports__per_safety_staff	Safety Reports Submitted per total staff	Safety Reports Submitted per dep	Safety Reports per miles	Safety Reports Submitted per fh	Number of Safety Reports Followed Up Feedback Provided	Safety Reports Followed Up per report submitted
Number of All Safety Related Events			Pos x2 3					Pos 2	Pos 2	
All Safety Related Events per dep			Pos 3		Pos 2		Pos 1	Pos 2	Pos 2	
All Safety Related Events per miles			Pos 1		Pos 1	Pos 1		Pos 1	Pos 1	
All Safety Related Events per fh			Pos x2 3		Pos 2	Pos 2	Pos 1		Pos x2 2	
All Safety Related Events per staff FTE			Pos 2	Pos 2	Pos 2	Pos 2	Pos 1	Pos 2	Pos 2	
Number of Occurences	Pos 1	Neg 1	Pos 3					Pos 2	Pos 2	Pos 2
Occurences per dep		Neg 1	Pos 3							
Occurences per miles										
Occurences per fh			Pos x2 3		Pos 2	Pos 2	Pos 1		Pos x2 2	
Occurences per staff FTE			Pos 2	Pos 2	Pos 2	Pos 2	Pos 1	Pos 2	Pos 2	
Number of Incidents		Neg 1								Pos 1
Incidents per dep		Neg 1	Pos 2							
Incidents per miles										
Incidents per fh										
Incidents per staff FTE										

	Reports_Followed_Up_per_safety_staff	Safety_Reports_Followed_Up_per_total_staff	Reports_Followed_Up_per_dep	Safety Reports Followed Up id per miles	Reports_Followed_Up_per_fh	Number of Hazards Identified from Sources Except Safety Reports	Hazards Identified per report submitted	Hazards Identified per dep	Hazards Identified per miles	Hazards Identified per safety staff
Number of All Safety Related Events			Pos x2 3		Pos 3					
All Safety Related Events per dep		Pos 2		Pos 1	Pos 3		Neg 1			
All Safety Related Events per miles		Pos 1	Pos 1		Pos 1		Neg 1			
All Safety Related Events per fh		Pos 2	Pos x2 3	Pos 1			Neg 1			
All Safety Related Events per staff FTE	Pos 2	Pos x2 2	Pos 2	Pos 1	Pos 2		Neg 1			
Number of Occurences	Pos 3		Pos x2 3		Pos 3					
Occurences per dep	Pos 3				Pos 3	Neg 2			Neg 1	Neg 1
Occurences per miles						Neg 1		Neg 1		Neg 1
Occurences per fh		Pos 2	Pos x3 3	Pos 1			Neg 1			
Occurences per staff FTE	Pos 2	Pos 2	Pos 2	Pos 1	Pos 2		Neg 1			
Number of Incidents	Pos 2		Pos 2							
Incidents per dep	Pos 2				Pos 2					
Incidents per miles										
Incidents per fh			Pos 2							
Incidents per staff FTE										
Number of Serious Incidents										
Serious Incidents per dep										
Serious Incidents per miles										
Serious Incidents per fh										

	Number of Total Risk Assessments Performed	Total Risk Assessments per safety staff	Number_of_Risk_Assessments_Initially_Rated_as_Low	Number_of_Low_Risks_in_the_Registry_after_assessment__mitigati	Number of Risk Assessments Initially Rated as Low	Number of Low Risks in the Registry after Assessment and Mitigation	Safety Reports Submitted per Flight Hour	Safety Reports Followed Up per Report submitted	Days for Implementing Decisions from Internal Safety	Number of Safety Meetings with External Organizations
Number of All Safety Related Events	Pos 3		Pos 2	Pos 2	Pos 2	Pos 2	Pos 1			
All Safety Related Events per dep										
All Safety Related Events per miles										
All Safety Related Events per fh	Pos 2		Pos 2	Pos 2				Pos 1		
All Safety Related Events per staff FTE										Neg 3
Number of Occurences	Pos 2		Pos 2	Pos 2			Pos 1	Pos 1		
Occurences per dep		Pos 2								
Occurences per miles		Pos 1								
Occurences per fh	Pos 2		Pos 2	Pos 2				Pos 1		
Occurences per staff FTE										
Number of Incidents										
Incidents per dep										
Incidents per miles										
Incidents per fh										
Incidents per staff FTE										
Number of Serious Incidents										
Serious Incidents per dep										
Serious Incidents per miles										
Serious Incidents per fh									Pos 1	Pos 1
Number of Accidents									Pos 1	Pos 1
Accidents per dep									Pos 1	Pos 1
Accidents per miles									Pos 1	Pos 1
Accidents per fh									Pos 1	Pos 1
Accidents per staff FTE										

Appendix 6: Significant Correlations between Operational Activity and Outcome Data

	Departures	Departures per staff	Miles Flown	Miles per Staff	Flight Hours	Flight Hours per Staff	Flight Hour per FTE	Flight Hours per Departure	Flight Hours per Aircraft	Ground movements
Number of All Safety Related Events		Pos 3	Pos 2	Pos 2	Pos x 2 5	Pos & Neg 4	Neg 3			Pos 1
All Safety Related Events per dep		Pos 3	Pos 2	Pos 2	Pos 4	Pos 3				
All Safety Related Events per miles		Pos 2	Pos 2	Pos 2	Pos 2	Pos 2				
All Safety Related Events per fh		Pos 3	Pos 2	Pos 2		Pos & Neg 4	Neg 3			
All Safety Related Events per staff FTE										
Number of Occurences	Pos 4				Pos x 2 4	Neg 3	Neg 3			
Occurences per dep					Pos 3					Pos 1
Occurences per miles										
Occurences per fh						Neg 3	Pos & Neg 3		Neg 3	
Occurences per staff FTE						Pos 3			Neg 3	
Number of Incidents	Pos 5	Pos 2	Pos 2	Pos 2	Pos x 2 4	Pos 3				
Incidents per dep		Pos 2	Pos 2	Pos 2	Pos x 2 3	Pos 2				Pos 1
Incidents per miles		Pos 2		Pos 2	Pos 2	Pos 2				
Incidents per fh		Pos 2	Pos 2	Pos 2		Pos 3				
Incidents per staff FTE										
Number of Serious Incidents	Pos 4	Pos 1		Pos 1	Pos 2	Pos 1		Neg 2		
Serious Incidents per dep		Pos 1		Pos 1	Pos 2	Pos 1				
Serious Incidents per miles	Pos 1	Pos 1		Pos 1	Pos 1	Pos 1		Neg 1		
Serious Incidents per fh	Pos 2	Pos 1		Pos 1		Pos 1		Neg 2		

Appendix 7: Significant Correlations between Demographic and Outcome Data

	Number of Company Staff	FTE Company	FTE Contractors	FTE Contractors per FTE Company Staff	Hours Flown per Pilot	Experience of Flight Crews Flight Hours	Experience of Ground Staff Years	Airrcaft Fleet	Aircraft Age Years
Number of All Safety Related Events	Pos & Neg 4	Pos 3			Neg 2		Pos 2		Pos & Neg 3
All Safety Related Events per dep	Neg 3								
All Safety Related Events per miles	Neg 2								
All Safety Related Events per fh	Pos & Neg 4	Pos 3		Neg 2	Pos & Neg 2		Pos 2		Pos 3
All Safety Related Events per staff FTE									
Number of Occurences	Pos x 2 3	Pos x 2 3	Pos 2		Neg 2		Pos 2	Pos 3	Pos 3
Occurences per dep						Neg 1	Neg 2		Neg 2
Occurences per miles						Neg 1	Neg 1		Neg 1
Occurences per fh	Pos 3	Pos 3			Neg 2		Pos 2		Pos 3
Occurences per staff FTE									
Number of Incidents	Neg 3								
Incidents per dep	Neg 2								
Incidents per miles	Neg 2								
Incidents per fh	Neg 3								