MASTER THESIS

Titel der Master Thesis / Title of the Master's Thesis

„The impact of safety performance indicators on operational safety of helicopter emergency medical service operations in Europe"

verfasst von / submitted by

Klaus Alexander Egger

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna 2020

Studienkennzahl lt. Studienblatt / Postgraduate programme code as it appears on the student record sheet:	UA 992 242
Universitätslehrgang lt. Studienblatt / Postgraduate programme as it appears on the student record sheet:	Risikoprävention und Katastrophenmanagement
Betreut von / Supervisor:	Dipl. Ing. Dr. techn. Gerald Lichtenegger

I. Declaration of originality

I hereby confirm that this assignment is my own work and that I have not sought or used inadmissible help of third parties to produce this work and that I have clearly referenced all sources used in the work.

Mondsee, August 15th 2020 Klaus Alexander Egger

II. Danksagung

Eine zusätzliche, langfristige Investition in Mitarbeiter ist nie selbstverständlich und daher möchte ich meinem Unternehmen, der ÖAMTC Flugrettung, für ihren immer großzügigen Umgang mit ihren Mitarbeitern danke. Allen voran meinem Vorgesetzen und Mentor Cpt. Reinhard Kraxner, der mir nicht nur dieses Studium ermöglichte, sondern mir immer motivierend mit Rat und Tat beiseite steht.

Bedanken möchte ich mich auch bei meinem Betreuer Dipl. Ing. Dr. tech. Gerald Lichtenegger für die Inspirationen, den Dozenten und allen voran dem Lehrgangsleiter Univ. -Prof. Dr. Thomas Glade. Gerade in Zeiten der globalen Krise, in der das freundschaftliche Netzwerk der Kommiliton*innen große Unterstützung bot, wurde mir klar, dass uns neben den Fachinhalten in diesem Studium vor allem eines gegeben wurde – eine gemeinsame Sprache!

Die Anforderungen der letzten Jahre und besonders des Jahres 2020 hätte ich nicht ohne meine Familie, meine Eltern, meine geliebte Frau und meine Kinder geschafft. Ich habe bei dieser Forschungsarbeit gelernt, wie wichtig es ist, neben einem Fundament aus Fakten, Erwartungen auszuschalten und trotzdem neue eigene Gedanken zuzulassen. Ich hoffe, dass meine Kinder diese Erfahrung auch machen dürfen und daher widme ich diese Arbeit meiner Frau und meinen Kindern: Danke Sandra, Valerie und Theo!

III. Zusammenfassung

Das Ziel dieser Studie ist die erstmalige Analyse der Verwendung von Safety Performance Indicators im Bereich der Flugrettung mit Hubschrauber.
Der Hintergrund der Arbeit ist eine derzeit noch unstrukturierte Anwendung von Indikatoren im Sicherheitsbereich. Wenn überhaupt werden Herangehensweisen aus der Linienluftfahrt kopiert, in der jedoch gänzlich andere Grundvoraussetzungen in Hinsicht auf valide Parameter gelten. In der einsatzorientieren Hubschrauberluftfahrt variieren Situationen sehr stark und somit müssen andere Parameter gefunden werden, um zuverlässige Messungen durchführen zu können, die schlussendlich ein valides Bild auf die Sicherheitssituation in einem Betrieb geben. Dazu wird in dieser Arbeit die Verwendung von Kategorien hinterfragt, welche zum überbetrieblichen Austausch verwendet werden können, um eine risikobasierte Einschätzung zu ermöglichen, und auch als Grundlage für ein mehrstufiges Modell dienen, das als Rahmenwerk für die Verwendung von Safety Performance Indicators gelten kann. Damit die weiteren Forschungsfragen wie Safety Performance Indicators kategorisiert und gemessen werden können und welche SPIs ein reales Abbild der Sicherheit in einem Betrieb darstellen, wurde neben einer Literaturstudie eine empirische Analyse mittels der Delphi Methode unter europäischen Rettungsflugbetreibern durchgeführt, anhand derer die aktuellen Herangehensweisen in Bezug auf die Verwendung von Indikatoren im Sicherheitsbereich erstmals erhoben wurde.
Die Ergebnisse bestätigen eine bislang rudimentäre Anwendung in der Branche und die Notwendigkeit zur Strukturierung, für welche Ansätze im obengenannten Modell gefunden werden können. Diese Arbeit ist für Sicherheitsexperten von Rettungsflugunternehmen interessant, da nicht nur mögliche Anwendungen, sondern auch für kleinere Organisationen umsetzbare Methoden zur Etablierung vorgestellt werden.

IV. Abstract

The aim of this study is the first-time analysis of the use of safety performance indicators in the field of helicopter emergency medical services.
The background of the work is a currently unstructured application of indicators in this industry. Approaches from the scheduled airline business are copied, notwithstanding that entirely different requirements apply with regard to valid parameters. In the mission-oriented helicopter aviation, the situation varies greatly and so other parameters have to be found in order to be able to carry out reliable measurements that ultimately give a valid picture of the safety situation in a company. For this purpose, the use of categories is questioned in this work, which can be used for inter-company exchange to enable a risk-based assessment and also serve as the basis for a multi-level model that can be used as a framework for the use of safety performance indicators. Further research questions such as, can safety performance indicators be categorized and by which means can they be measured, as well as which SPIs represent a real image of safety in a company, are challenged. In addition to a literature study, an empirical analysis using the Delphi method was carried out among European rescue flight operators to question the current approaches in regards to the use of indicators in their respective safety management system.
The results confirm a rudimentary application of SPIs in the industry and the need for structuring, for which approaches can be found in the represented model. This work is interesting for security experts from HEMS operations, because not only possible applications but also methods for establishment that can be implemented in smaller organizations are presented.

V. Table of content

VI. Table of figures

VII. List of tables

VIII. List of abbreviations

APF ....Aviation Performance Factor
ASC-IT ....Aviation Safety Culture Inquiry Tool
BBN...Bayesian Believe Networks
EASA..European Aviation Safety Agency
ECCAIRSEuropean Coordination Centre for Accident and Incident Reporting Systems
EHA...European Helicopter Association
EHAC ....European HEMS & Air Ambulance Committee e.V.
ERP...Enterprise Resource Programs
ESD.. Event Sequence Diagrams
EU OPS...Air Operations Regulation, VO (EU) 965/2012
FAA .... Federal Aviation Administration
FDM ....Flight Data Management
FRAM ....Functual Resonance Analysis Method
HAI ....Helicopter Association International
HCI ....Hybrid Causal Logic
HEC...Human External Cargo
HEMS ....Helicopter Emergency Medical Service
HLI.. High Level Indicator
ICAO ....International Civil Aviation Organization
IFR .... Instrumental Flight Rules
ISO ....International Organization for Standardization
LLI .... Low Level Indicator
NLR ....Netherlands Aerospace Centre
OECD .... Organisation for Economic Cooperation and Development
OGP .... Oil and Gas Producers Standards
PRA... Probabilistic Risk Assessments
QMS ....Quality Management System
RIF ....Risk Influence Factor Model
SME .... Small or Medium Size Enterprises
SMS ....Safety Management System
SPI ....Safety Performance Indicators
UN .... United Nations
WANO ....World Association of Nuclear Operators

1. Introduction

1.1. Motivation

Aviation is one of the safest means of transport. This widespread opinion depends strongly on the parameters of the comparison – in this case it refers to the distance covered. Comparing absolute numbers in 2018, 523 persons died in airline accidents worldwide (cf. IATA, 2019: 6) related to over 25.100 fatalities in road accidents only within the European Union (cf. EC, 2019: 1). If fatalities are compared with the total number of transported passenger's, railway systems are beneficially leading these statistics (see ERA, 2019).

The bias for measuring and controlling safety performance spreads, if a closer look into the different categories of the aviation industry is taken. Most references are made to the commercial airline business which is dominated by a very restrictive, highly standardized environment that leaves little room for individual decisions. Safety as a business management process and legal requirements in aviation are aligned to the airline industry. But on the opposite side of the spectrum there are military or police helicopter operations or the helicopter emergency medical service (HEMS) where situations are very dynamic and factors of influence vary with almost every mission. This fact leaves the sector of the helicopter industry with the task of finding standardized safety monitoring methods which are comparable for advanced cross company risk analysis, give foundations to communicate safety performance within the management and the company and still give a plain, self-reflected picture of the safety culture in an operation.

Traditional methods to evaluate operational safety risks were mostly based on tracking so called trailing indicators, like lost work days, numbers of accidents and incidents and therefore represent something like an after-the-fact hazard detection. In order to reproduce the influence of a safety culture as well as the effects of proactive safety measures, an operationalization of safety performance indicators is essential.

In terms of oversight aviation, authorities are moving away from their traditional form of supervising companies' performance and transfer to a focus on inspecting the companies compliance and safety management (cf. Alfred Roelen & Klompstra, 2012: 1). Unfortunately it still appears in the rescue helicopter industry that the same standards and scales are applied which are normally used for scheduled airlines or helicopters in the oil business where specific standards of oil and gas producers (OGP) are held.

Parameters used by scheduled airlines or oil business helicopters are often flight times averaging far above one hour or for instance parameters for landings are very predictable as the airport or oilfield are always the same. In comparison with the air ambulance system in central Europe the average flight time having a passenger on board in the fleet of the ÖAMTC Flugrettung for example, the biggest Austrian helicopter air rescue provider is 8.2 minutes. None of the nearly 20.000 missions in average flown per year by this fleet are having the same destination (see ÖAMTC Flugrettung, 2020).

Also data gathering methods like flight data monitoring need to be viewed from a different point of view. Non standardized flight paths create different challenges in the utilization of the gathered data.

In this thesis I would like to examine the specific way of verification of safety through Safety Performance Indicators (SPI). *“Safety management is more a regulatory requirement than an industry best practice”* claim Roelen and Klompstra (2012), but regardless the jurisdiction it should be the ultimate goal of all safety efforts to prevent harm to persons or property. This harm is connected to the safety performance of an operation and its ability to learn lessons from occurrences and its resilience to react to incidents or accidents (cf. (Gerbec, 2013) and (Øien, Utne, & Herrera, 2011)). Roelen and Klompstra (2012) also state that *“…As of yet, the operational experience with measuring the effectiveness of safety management systems is very limited and there are many questions yet to be answered on measuring safety performance, demonstrating compliance with safety management regulations…”.* The statement in this study (ibid) that many states are moving away from an active role of supervision due to a high amount of inspectors required might be challenged as the understanding about the authorities about the diversity of operations in the aviation business might also lead to more oversight activity than supervision through the states. Aviation can be seen as an environment of high reliability organizations (HRO) where real accidents declined to a rate of 10^{-6} or less and the chance to learn from accidents is little (Amalberti, 2006: 270). Meaningful indicators bear the opportunity to observe changes in safety and risk expositions without suffering the immediate effects. This thesis would like to illustrate a reference framework of the usage of the safety performance indicators in the environment of mission based helicopter aviation and to research the link between outcome parameter and the effectiveness of safety management systems.

1.2. Starting point, objective and structure

The baseline problem is an unstructured utilization of safety performance indicators in the environment of helicopter emergency medical service. SPIs are an important tool in the world of aviation to monitor safety. Furthermore it is statutory by current European regulations (EC, Regulation 965/2012: AMC1 ORO.GEN.200). This legal requirement induces helicopter companies to use such indicators, but it is assumed that most of the few used indicators are neither significant nor purposeful for the original demand – proactively measuring safety.
The above mentioned, very controlled environment in the fixed wing airline world represents a complete different regime and often reproduced indicators used in another setup like the helicopter aviation which leads into the void. Based on a rather small part of the industry, HEMS actually represents merely one-sixth of the global helicopter market (cf. Seymore, 2018: 4), and a just slowly progressing build-up of knowledge, there is factually no literature dealing with the concepts of safety performance monitoring in the field of helicopters or even continuing further to emergency medical service.

The setup of this paper is therefore divided into a theoretical part and an empirical part where inputs from the helicopter aviation industry in the medical emergency sector throughout Europe are processed using the Delphi analysis. The hypothesis laid down in the second chapter will guide through the research made and are challenged in the topic analysis. Notwithstanding the hypothesis used, this thesis shows a mentionable amount of explorative research in the analysis section due to the above-mentioned fact that the content of this field of research is currently highly unstructured. The explorative approaches should help to gain insight into the needs and possible ways of applications of safety monitoring, utilizing safety performance indicators. Therefore, it has been tried to attempt a structured system consideration in the chapters 4.1 and 4.2 providing a systematic overview and the context boundaries. Chapter four merges the knowledge gained from the theoretical part and the empirical part which is described in detail in chapter five. A reflection of the findings in the context with the current status, as well as possible further fields of research and applications is covered in chapter six. Chapter seven, conclusion, should provide a summary of the previous chapters, and implications for the HEMS industry are made.

1.3. Definitions and basic principles

Every industry has its specialized terminology. In this particular case of safety monitoring, the different aviation industries create, in a progressive development, their own way and definitions which are not yet universally made, or are interpreted in different ways. Relevant terms are listed below in a thematic order.

International Civil Aviation Organization
The International Civil Aviation Organization (ICAO) is a specialized agency of the United Nations (UN). It´s aim is to reach consensus between the 193 member states on international civil aviation Standards and Recommended Practices (SARPs) and policies in support of a safe, efficient, secure, economically sustainable and environmentally responsible civil aviation sector. ICAO claims in the respective annex to the convention on international civil aviation, known as the Chicago Convention - (AIZ, 1971), that the major factors like the level of acceptable risk, have not been universally established. (cf. ICAO, 2016: 3-4)

ICAO Annex 19 – Safety Management
On the regulatory side there is the ICAO with the aim to regulate air traffic on an international level. ICAO is also developing safety standards, and the established standards are binding for the member countries. (see ICAO, 1944)
Attached to the Chicago Convention are annexes which have to be separately ratified by the member countries. The ICAO Annex 19 – “Safety Management” to this convention addresses aviation safety management under the provisions: Safety Management System (SMS) framework, State Safety Program (SSP) framework - Safety Policy and Objectives, Risk Management, Safety Assurance and Safety

Promotion as well as State Safety Oversight System and Safety Data Collection Analysis (cf. ICAO, 2016: 3-1 et seqq.).

High Reliability Organizations

HROs are organizations facing constant exposition to serious and complex risks and nevertheless succeed in avoiding catastrophic accidents (see Roberts, 1989). Finding its origins in the safety management of nuclear aircraft carriers, the term nowadays includes industries like the chemical industry, oil and gas explorations or nuclear power plants. Those organizations are looking for single events that show evidence of latent conditions (see Hopkins, 2009). Even though the aviation business and the aforementioned industries are already utilizing safety performance indicators and can therefore be counted as HRO, the helicopter emergency medical service is not applying them on a systematic basis. Indicators for nuclear power plants for instance are core damage frequency (CDF), radioactive release frequencies or system safety unavailability. Those indicators are used for a benchmarking within the nuclear power industry on a worldwide basis, developed by the World Association of Nuclear Operators (WANO) (cf. Øien et al., 2011: 153). The structured system in these high reliability organizations represents an important approach in safety science.

European Aviation Safety Agency

The methodologies given by the ICAO have to be implemented into national law which again led to different interpretations in the past. Nevertheless the states were responsible for their own flight safety. Upon decision in the year 2002 by the European Parliament and the council of the European Union, the European Aviation Safety Agency (EASA) was founded (EASA, 2020a). EASA's mission is to harmonize the different aviation relevant laws of the member states and to implement high level uniform safety and environment standards. The relevant legislation for flight operations containing safety pertinent regulations is the so called EU OPS which stands for Air Operations Regulation (see EC, Regulation 965/2012). The section ORO.GEN.200 as amended by Sept. 2019 in the EU OPS describes the provisions for a management system in a flight operation. Main focus is put on compliance and safety management systems. The rule makes a difference between complex and non-complex operator and especially for complex operators the requirements are described for hazard identification processes, risk assessment and mitigation processes internal safety investigations, emergency response planning, continuous improvement, management of change and safety performance monitoring and measurement. The paragraph for safety performance monitoring and measurement outlines that: *"Safety performance monitoring and measurement should be the process by which the safety performance of the operator is verified in comparison to the safety policy and objectives."* (EC, Regulation 965/2012: AMC1 ORO.GEN.200). With setting these goals, the regulator determines the obligatory use of safety performance indicators.

Safety Performance Indicators

Performance indicators have a long history in business administrations as a method for evaluations. In the field of safety, indices were used already before 1980, but named differently (cf. Kjellén, 2009: 486). In aviation the first lead found was in the 1990s with the implementation of JAR-OPS, a precursor of the now effective European regulations.

Those keys, ratios or indices are today called safety performance indicators and are an interdisciplinary matter in the areas of social science, engineering, natural and statistical science. As mentioned before, there are various undetermined classifications, but in general all attempts to group them resulted in the common understanding that they either are leading or lagging indication. With the term “leading” indicators are those characteristics meant, which indicate the level of proactive measurements. OECD, the Organization for Economic Cooperation and Development, refers to those indicators as activity indicators (OECD, 2014: 5). For instance, the quantity or extent of proactive safety training or safety audits held in a flight operation. On the other side of the scale are lagging indicators subsuming all indices that are historical and accordingly measuring the outcomes and occurrences. (see Kasper, Piric, de Boer, Roelen, & Karanikas, 2019)

Most of the currently used indicators represent lagging indicators and correspond to an after-the-effect hazard detection. Remarkably almost every publication of the aviation authority is based on outcomes (EASA, 2020b).

Knut Øien defines “An indicator is a measurable/operational variable that can be used to describe the condition of a broader phenomenon or aspect of reality” (cf. Øien, 2001: 6) based on a combination of (Hellevik, 1999) and (Gray & Wiedemann, 1999).

A sound definition of the interferences between the terms safety indicator and risk indicator is important, as they are sometimes used interchangeable. Depending on what should be measured, the aspect could reflect a safety process, and then it would be measured by a safety performance indicator. If it reflects a factor influencing risks, the indicator needs to be included in a risk model and is then called risk indicator (see Figure 1). The risk model could be an incident based approach or a probabilistic risk assessment and the indicator is consequently linked on a causal connection while the safety performance indicator is based on assumed relations or the use of correlation. (cf. Øien et al., 2011: 149-159)

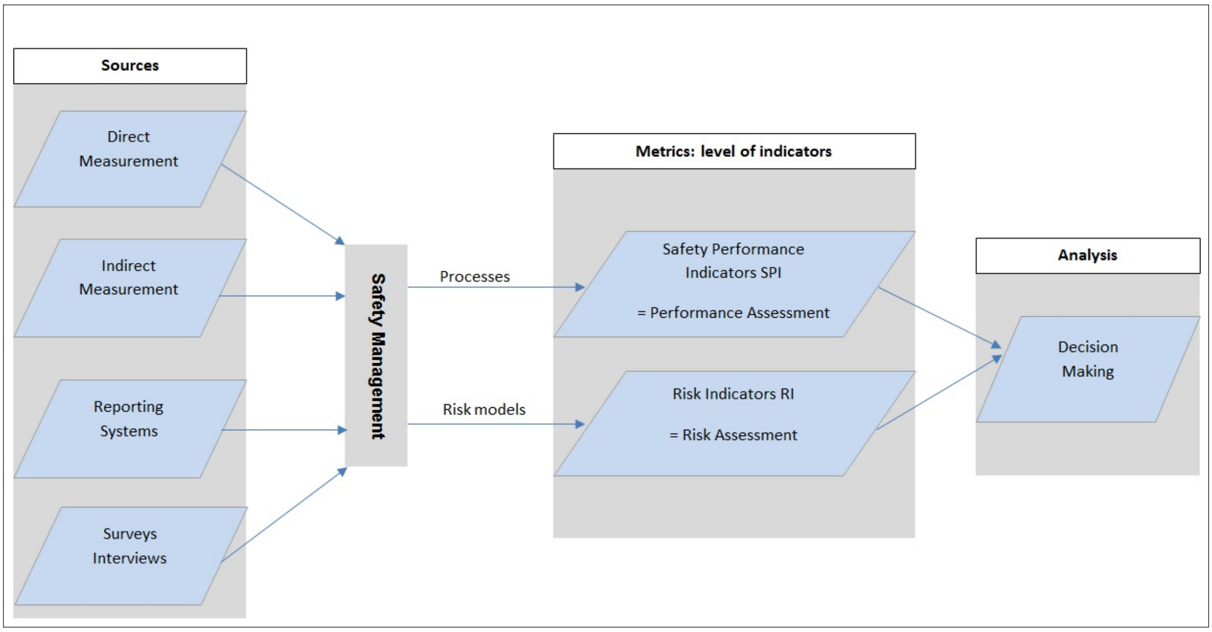


Figure 1: Sequence of safety information processing with focus on safety performance indicators (own representation)

The classification of safety performance indicators is one key to create a common basis to compare safety metrics in order to apply a risk based approach and will be reviewed in chapter four of this paper.

Safety Performance Targets

Safety performance targets are usually referred to as safety objectives or safety goals. Technically the difference lies in the time frame as goals are intended for a longer time period and may not be quantifiable while it is a requirement for the short term objectives to be always measurable (cf. Pierce, 1995: 125). There is evidence in the literature about finding security targets; however it is important to mention, how cruicial it is, that SPIs are tied to the organization's safety goals and objectives. In analysis of safety performance programs, discontinuity between the focus of activities of an organization and the performance measures have been identified (cf. Janicak, 2015: 26). Safety performance targets can be process or outcome-oriented. An example of a process-oriented safety objective would be the increase of safety reporting levels versus the reduction of annual numbers of third party damages represent an outcome-oriented value in HEMS operations (cf. ICAO, 2009: 4-4).

Safety

Discussing the way of measuring safety, the definition of safety itself should be clarified. While different engineering standards and codes tried to describe the term in detail, Karl Weick from the University of Michigan accomplished the shortest accurate description of safety: *"it´s a dynamic non-event"* (Weick, 1987: 112). With dynamic is meant that the outcome of the non-event cannot be guaranteed (Hollnagel, 2014: 9). A noticeable fact is that most of the international standards found, like ISO 31000 which talks about risk management, ISO 9001 which talks about quality management or ISO 45001 describing occupational health and safety assessment or patient safety standards like the Luxembourg declaration on patient safety or the London protocol do talk about safety but without giving a clear definition.

Even the binding European regulatory framework does not provide an applicable definition of safety or what the aim of establishing the stipulated safety is.
Compared to other industries, aviation is historically one of the most exposed to public safety considerations. Usually passengers in the recent past paid at lot of attention to the reputation of an airline or specific aircrafts and even accepted higher rates for tickets in order to fly with a seemingly safe airline. Airlines with an unfavorable accident statistic were, depending on the geographic culture area, not competitive. This established market-orientated procedures, which helped the process of identifying risks and confine them to a tolerable level. (cf. Mickel, 2011: 6)
The point of view that there is a residual risk which needs to be minimized according to market needs and defines the risk appetite of an organization as well as ethical principles is a basic understanding of the environment of high reliability organizations.
Nevertheless, one of the few definitions found for example in ISO/IEC guide 51:2014 which says that safety is *"…freedom from unacceptable risk…"* (ISO, 2014: 4) is conform to ICAO's view where acceptable risk levels have to be achieved (ICAO, 2016: 1-2). Another definition by ICAO says that safety is: "…the state in which the possibility of harm to persons or of property damage is reduced to, and maintained at or below, an acceptable level through a continuing process of hazard identification and safety risk management" (cf. ICAO, 2009: 2-1). Nevertheless, those definitions include predefined terms, which will be analyzed in this paper and therefore the definition of safety as used in this paper is the following:

"A condition of certainty for an individual or system to be highest possibly protected from unjustifiable exposure to adverse conditions"

This definition also reflects the in aviation conventionally used ALARP principle. ALARP stands for "as low as reasonably practicable" and reflects the approach residual risk is imminent and should be reduced as far as realistically practicable (cf. Bai & Jin, 2016: 719-720). Having this definition of safety determined, the next step would be to describe the safety goal, thus for an individual organization to identify what risks are relevant and what is considered as acceptable and what as unacceptable. To define what is acceptable, levels need to be defined which again determine the need to measure. For every well-grounded analysis of performance measures, it is a central requirement to provide regular, valid data (cf. Hatry, 2006: 13). However, there appear to be differences between common performance measures and safety metrics. Christopher Janicak describes in his book "Safety Metrics" the difference between performance measures and safety metrics as safety metrics are the standards of measurements while performance measures evaluate the safety process, so a series of safety metrics compile to a safety performance measure (cf. Janicak, 2015: 6).

The above mentioned terms fit together to a framework of how safety is measured in an organization. There is several evidence in the literature about frameworks, the one which provides the widest options of application, even though it was originally thought for the chemical industry, is the one made by the OECD (OECD, 2014: 11-35). Christopher Janicak's presentation (see figure 2) of a framework depicts the hierarchical correlation of the processes including OECD's philosophies (cf.Janicak, 2015: 7).

Figure 2: Safety Performance Framework (Janicak 2015: 7)

Risk

Risk is commonly described as a function of probability of harm and severity of that harm (ISO, 2014: 4). The prefix "unacceptable" is vague and dependent on the culture of the user. What can be acceptable to one individual or system, might not count for another. The probabilities of harm can be judged by the quantity of past events or based on judgement of experts (cf. Duijm, 2015: 25). Estimations of the possible severity of harm are made through maximum-credible accident scenarios (see Khan, 2001) based on extrapolation of known events and assumptions. Limitation of this method is the combination of several factors which can contribute to top-level accidents (cf. Kasper et al., 2019: 4).

Safety culture

The level of risks accepted is part of the safety culture in an organization. The safety culture is an often used term in this paper, as it not only describes the level of accepted risk by an organization and its individuals, but also the following six culture characteristics: commitment, justness, information, awareness, adaptively and behavior (see Montijn & De Jong, 2006). Those six characteristics mean to describe the values of the stakeholders in terms of how they are dedicated to safety, how they reflect their behavior of safety, how they communicate about safety, how their alertness and attentiveness is, how they are able to adapt to new situations concerning safety and how their behavior is considering safety activities.

Safety Management System (SMS)

All comes together in the safety management system. It represents the framework of an organization where safety policies depict the safety goals and the desired safety culture, utilizing processes and safety objectives with the aim to provide a structured management approach. (cf. ICAO, 2009: 9-1)

Incidents / Accidents

Additionally to the ICAO definitions which focus on the operation of an aircraft, in this paper all occurrences in association with the operation itself are seen as accidents or incidents. This means events in conjunction with the handling of a patient or technical rescue operations are included. While an accident is defined by a person being fatally or seriously injured or an aircraft sustaining damage or structural failure, an incident means any occurrence not being defined as accident.

2. Hypothesis, research questions and objective

Safety Management and the related challenges are in a supposedly safe environment like the western world a topic which is getting a lot of attention. While static processes are rather easy to define and monitor, it is an interesting challenge to try to determine parameters of safety in dynamic systems with lots of variables. The underlying question is: What is considered as a safe system, under which parameters do the set definitions apply and how many variabilities can a system take to be considered safe.

The helicopter emergency medical services represent a small specialized field in the world of aviation including implementations in ambulance systems and therefore political values as well as including the whole sphere of patient safety. Although the medical aspects are excluded from this paper it appears as if this regime deals with a similar dynamic - especially from a preclinical point of view.

The ICAO recognized the importance of the organizational influence and the comparison with the condition in the whole industry (Figure 11). Events, for instance organizational changes, can have a great influence on the safety performance: "S*uch deep organizational changes in a company which uses hazardous technology can reduce the effectiveness of its process safety management system unless the implications of change are consciously and carefully addressed*" (cf. Baram, 1998: 191). Therefore it is necessary to get a holistic picture inside and outside an operation. The contexts about which methods to measure safety with show a real picture of safety and how they could be compared with industry standards, leads to safety performance indicators. The research questions aim to find out how SPIs are currently used in HEMS companies and if, with the current approaches, actual safety performance can be measured.

Research questions

1) Hypothesis: Safety performance indicators can be classified in applicable categories.

    a) What are relevant categories of SPIs in helicopter emergency medical services?
    b) Which SPIs can be monitored by the same means?

2) Hypothesis: Safety performance indicators show an impact on operational safety.

    a) Which SPIs reproduce a real picture of the safety situation in an HEMS operation?

3) Hypothesis: Safety performance indicators have to be found specifically for each operation and still have to be comparable on an industry-wide basis.

    a) What are the challenges to define SPIs for a specific operation?
    b) How do SPIs remain comparable to associate cross-company performance?

It’s the aim of the first hypothesis to ascertain the challenges of defining feasible safety performance indicators and the way to measure them. Different indicators measure different facts. For instance, indicators to measure a culture in an organization are presumably not capable of measuring a certain exposure to risks on a numerical scale. The biggest discrepancy can be found in the difference of the basic classification of reactive and proactive measuring of safety performance. Therefore, the application of categorizations and their effects on the measurement of safety should be questioned.

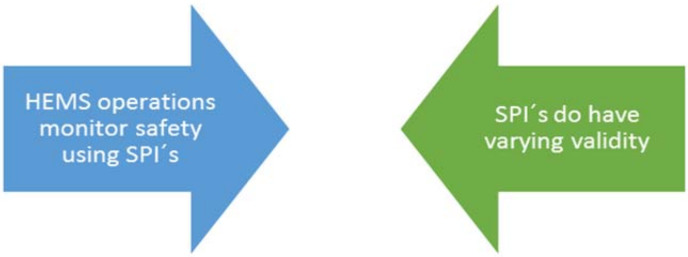


Figure 3: triggering process for this thesis
(own representation)

The second hypothesis describes the common practice of using fatal accident counts as an indicator of safety. In an industry which consists of high reliability organizations, accident rates are extremely low and the statistical value of counting accidents is marginal. So-called “black swan events”, an event which has not happened in the aviation industry before, could have a dramatic impact on these statistics without having an actual validity on the safety in an organization. Complex and diverse forms of organizations tackle the challenge to adopt performance indicators for the specific operation assessing a valid statement of safety.

These specific indicators could presumably be merged into categories to be comparable again with industry standards and benchmarks in order the keep the holistic view. Diametrically opposed operational areas make it hard to apply the same measurements on different operations. What are the challenges in defining specific SPIs and how can they still be compared, is therefore the substance of the third hypothesis.

3. Scientific methods and survey

3.1. Scientific methods used

For the theoretical part of this paper a review and synthesis of literature has been applied. As the topic of this thesis represents a very small field in aviation, where almost no research takes place, there are only a few specific articles available. Hale already described this deficiency as “a black hole in research and literature”(cf. Hale, Baram, & Hovden, 1998: 11). No papers have been found engaging in research on the subject of safety performance indicators concerning HEMS operations (compare with Table 1). The survey of this thesis showed that only half of the safety experts of the responding companies have heard of newer analysis methods (like STAMP, FRAM, RIF or HFACS); hence the topic of SPIs in helicopter operations seems to be rather new to the world of the helicopter industry. The review focused on studies about high risk industries and journal reports like contributions to the journal of air transport management or studies especially those carried out by the National Aerospace Laboratory NLR of the Air Traffic Safety Institute in Amsterdam in the Netherlands. In the context of the oil and gas business in northern Europe several studies were made with the focus on accidents in helicopter offshore operations. The first studies investigating performance indicators for aviation safety were found from the year 2006, published in the journal of air transport management.

Structured literature research was applied through finding specialized books. From the authors found in these technical books a forward looking literature research was conducted. Most parts of the literature found focused on probabilistic methods for mere technical systems like the chemical industry or the nuclear power industry. With the knowledge that estimated 80% of failures in aviation can be dedicated to human errors (cf. Shappell & Wiegmann, 2001: 59), the approach of calculating component failures seems not to be adequate. The criterion for the aviation relevant academic literature used was primarily the relevance to the topic and secondly the date of publication.

For the subjects concerning scientific methods, a rearward looking literature research was made. The Delphi Analysis was selected as method for the survey out of 32 different survey methods. Research about the methodology was made using technical books about organization research and scientific methods.

For the preparation of the survey questions, initially the so-called facet theory was chosen. The facet theory helps to find a structured way to generate tuples for the questionnaires (see (Häder, 2014) and (Ammon, 2009)). Unfortunately, the facet theory proved to be an excellent tool in the field of social science, but would not be the perfect tool for technical approaches favored in the present survey ((see Borg, 1996) and (Canter, 1985)). The basic principles to generate the survey were found in the literature of (Hug, Poscheschnik, Lederer, & Perzy, 2014) and (Faulbaum, Prüfer, & Rexroth, 2009).

Sources for the literature research were scientific search portals like ScienceDirekt or google scholar. Predominately used was the scientific library of the university of Vienna and the online service “u:search”. Legislative documents and texts where downloaded in the available version in effect from the European Law database EUR-Lex.

The primary literature listed below represents the most important technical literature and papers in the field of safety performance indicators.

Table 1: overview of the most important primary literature used

Background	specification of the paper	Author	Classification of the literature
safety research	SOL – Safety through organizational learning: A method for event analysis	Fahlbruch, Schöbel (2011)	indirect connection
	Safety metrics: Tools and techniques for measuring safety performance	Janicak (2015)	direct connection
	Building Safety indicators: Part 1 – Theoretical foundation	Øien (2011)	direct connection
	Managing the risks of organizational accidents	Reason (2008)	recognizable connection
aviation safety research	How does aviation industry measure safety performance? Current practice and limitations	Kaspers et al. (2019)	direct connection
	The art of measuring nothing: The paradox of measuring safety in a changing civil aviation industry using traditional safety metrics	Lofquist (2010)	direct connection
	The challenges in defining aviation safety performance indicators	Roelen and Klompstra (2012)	direct connection
	Effectiveness of risk controls as indicator of safety performance	Roelen et al. (2018)	direct connection
	Proposing safety performance indicators for helicopter offshore on the Norwegian Continental Shelf	(Herrera, Hollnagel, & Håbrekke, 2010)	direct connection
	Accident models and organizational factors in air transport: The need for multi-method models	Roelen et al. (2011)	indirect connection
Resilience engineering	Optimum System Safety and Optimum System Resilience: Agonistic or Antagonistic Concepts?	René Amalberti (2006)	recognizable connection
High reliability organizations	Human Performance Enhancement in High-Risk Environments	O´Conner and Cohn (2010)	recognizable connection

3.2. Delphi Analysis

Referring to the ancient oracle of Delphi a secret study for planning strategic weapon systems, made by the US military founded RAND Corporation in the fifties, was called “Project Delphi”. This established the name for a methodology to find consensus on a structured way between experts (cf. Kühl, Strodtholz, & Taffertshofer, 2009: 458). The original idea was to find a controlled way to question specialists and return the gathered, anonymized answers in several stages back to the panel of experts to find consensus while at the same time excluding the influence of psychological and situational factors like persuasion, antipathy or the influence of building majorities in groups. The multi-stage approach gives the chance to – after learning anonymously about the group opinion – rethink and revise the assessment of the topic. (see Ammon, 2009)

Michael Häder (cf. Häder, 2014: 30) finds four ways of Delphi analysis methods depending on the searched aspects:

1. Delphi-analysis to aggregate ideas
2. An analysis to predict diffuse circumstances
3. Delphi-analysis for identification and qualification of an expert-opinion to a certain matter
4. Delphi-analysis to find consensus among experts

In this study the diversity of interpretation in a relatively new field and questions to the aspects of helicopter aviation should be answered. It is the aim of the Delphi-analysis to identify pertinent issues and spot field requiring action by questioning the experts of safety in the European helicopter emergency medical services which corresponds to the third item in Häders analysis (ibid).

In principal, the Delphi-analysis is a quantitative research method. Still there are some aspects of combining qualitative and quantitative research methods in organization research (cf. Kühl et al., 2009: 458 et seqq.) which made it the appropriate tool for this paper. The decision to use a combination of qualitative and quantitative research methods is justified in the topic of this thesis: There are barely any reliable sources for numerical data and therefore research work is necessary to find valid categories with comparable categories and variables. This first round in a two-tier process is supposed to generate the precondition to create measurement instruments and the opportunity to investigate dependencies of variables. In the second questioning the given anonymized answers will be summarized and given back to the expert group and challenged.

The anonymity is a critical item in Delphi Analysis. It is this anonymity of the individual responses and the anonymity of the survey participants which should guarantee that the respondent is free of any group dynamic effects. In this survey the utilization of the safety managers of the leading HEMS operators as an expert group is used to reflect the compiled outcome as well as determine and qualify their opinions on the current usage of SPIs in the industry.

There is no common understanding in the literature about how big the panel of experts should be. Fred Woudenberg (1991) concludes in his study “An evaluation of

Delphi" that: *"A statistical aggregate of several individual judgements is more accurate than judgement of a random individual […]"* thus he describes three persons as not enough due to the fact that accuracy of the method may suffer. (cf. Woudenberg, 1991: 137).
Häder describes in his book 11 different scientific literatures describing the size of the expert panel in a Delphi-analysis. The numbers of experts range in these studies from three to over a couple thousand persons.
For the survey of this study the expert panel compounds of a group of high level experts which represents the safety personnel which is operating in the context of SPIs. As all participating respondents are currently safety managers in HEMS operations throughout Europe, the answers provided can be valued as well-founded. There is only a small group of people who are theoretically eligible to answer questions about this topic. This is why principals which are normally intended to be fulfilled in these kinds of studies have been ignored (e.g. a certain demographic distribution or a certain gender ratio). As the questions are rather related to a specific technical and organizational field than to personal opinions, demographic distributions are not relevant.
The Delphi-analysis purpose is not meant to archive consensus between the experts but to find the lowest common denominator and the current level of application in the regime of performance indicators. Consequently a method with repeated questionnaires and feedback from the previous round necessitates the experts to reconsider their requirements. Due to advantages in cost and workflow the survey was held using the online questioning tool "Questback". Originally planned face-to-face interviews have been canceled due to the corona pandemic crisis during the research period of this paper in 2020.

Pretest and set up of the questions

To check the function and usability of the survey, a pretest was made utilizing as an expert group four members of the compliance and safety management department of the ÖAMTC air ambulance. According to the pretest objectives given by Häder (2014) the pretest group was questioned if:

- the questions were understandable
- the respondent had any problems understanding the task
- the questionnaire was too long or too short (e.g. the interest of the respondent could be kept up throughout the interrogation)
- any questions raised discomfort (e.g. through talking about confidential information)
- any technical problems arose

The author of the study used the pretest to check:

- the average time to answer the survey (e.g. if the afforded time is reasonable)
- sufficient distribution in subjective answers was given (excluding the answers corresponding to their company as they work all for the same organization)
- the cover letter motivated and instructed sufficiently

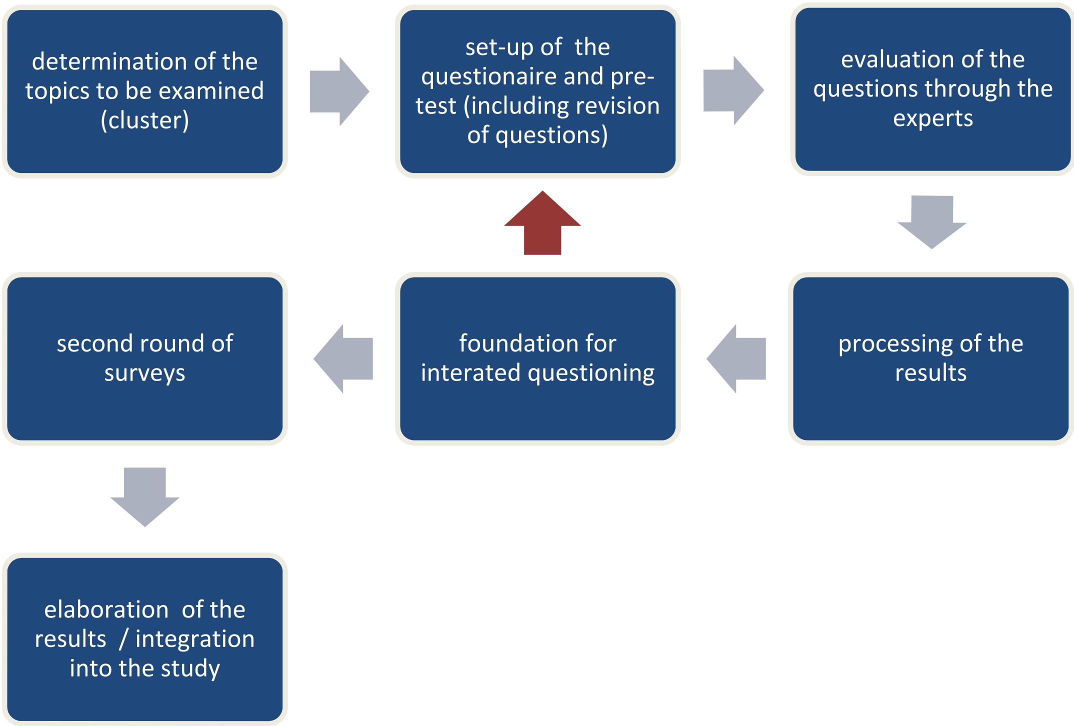


Figure 4: Sequence of process steps for the Delphi analysis (own representation)

For the pretest five persons with diverse occupations in the professional work environment of safety and compliance were chosen to assess the test set-up.

In general, all test persons responded that the questions were understandable. Also were there no problems understanding the task, with the exception that one respondent needed additional information to be able to answer the questions and two respondents had to guess on company data as they did not have the information on hand (which was anyway not the aim of the pretest). The time necessary to answer the questionnaire was experienced as not too long and no discomfort was raised while answering (e.g. for sharing confidential information). No technical problems or complaints were found which were in the sphere of the online questionnaire. One suggestion of the test group was to add a possible answer for the case if no answer could be given. This recommendation was not taken up, as an accumulation of answers with no context to the content, due to carelessness should be foreclosed. None of the questions are mandatory, so if one of the respondents could not give an answer, this question could be left out.

In the pretest phase some questions were changed with regards to the content because some questions turned out to be leading questions. Questions asking for company data which is used to compare identified approaches with company structures were changed as to refer to a specific time period e.g. "How many helicopters are currently in your operation?" or "How many flight missions does the operation you work for achieve on average during one year?". In order to get clear, assignable answers, some questions were split up into two questions.

The questions that were asked were categorized in several clusters to structure the processing of the responses. The queries and the cluster do not directly correspond to the research questions. Direct questions about the content of the hypothesis would in some parts require in-depth knowledge of the application on SPIs which is not assumed.

The first cluster provides information which helps with the valuation of the information obtained in the whole questionnaire. It is about operations parameters and management approaches which are compared with the data concerning the application of SPIs.

The second cluster of questions aims to find information in the regime of the first hypothesis. It should be found out which categorizations of SPIs would be beneficial for organizations and whether the categorizations offer the conditions for comparison at all.

In the third cluster the categorizations and impact on operations are the research objective. At the end of each survey, a field to leave comments was included with the intention to give a chance for feedback and also to give input to the topic. No responses contributing to the theoretical content of the study were found in the feedback.

The clusters were made a-priori to structure the questions and remained the same for the second round of the survey while the content of the questions of the second round was inducted by the answers of the first round.

The survey included 35 operations providing helicopter medical emergency services in Europe. In order to reach all companies in Europe, online research and existing contacts throughout Europe were used. In the first stage, 11 companies replied, which means a return rate of 37%. More than half of the responders came from DACH countries but also UK, Luxembourg, Norway and Hungary based operations answered. No response came from countries like France, Spain or Italy which possibly can be seen under the context that during the time of the survey the corona crisis was on a first peak across Europe and, of course, demanded a lot of effort from safety personnel of all kinds of preclinical care organizations. In general, and especially under these exceptional conditions, a response rate of 37% can be considered as very high and thus also shows that it is an issue interest. This can also be confirmed by the fact that in the second round of the survey a favorable return rate of 82% was reached.

In a small industry it is often difficult to get reliable data and since the number of samples is very small, the high result of the return rate of the second round is even more beneficial as it accordingly contributes to the usability of the data. The range of organizations which took part in the survey, spans from small companies running two helicopters up to governmental organizations and international companies with up to ninety aircrafts. The number of employees behaves accordingly, which is relevant for a comparison of cultural aspects in the different sizes of organizations.

4. Analysis

4.1. Current approach to safety monitoring

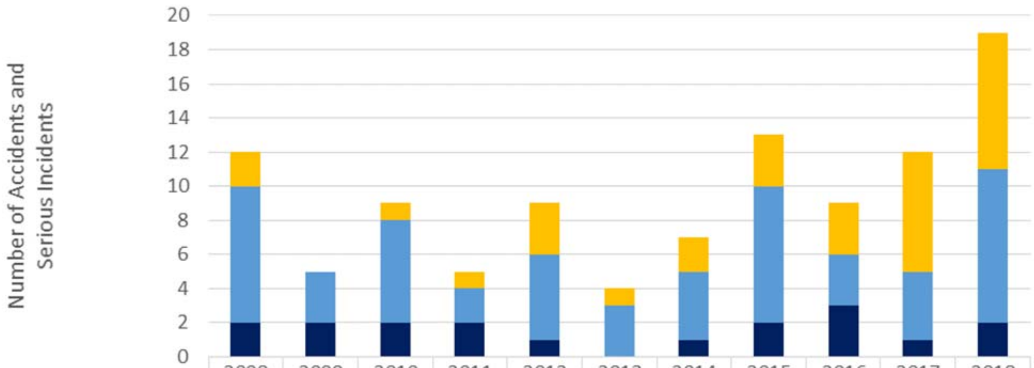
Organizational learning and aspects like human factors or organizational and system aspects gained influence over the last thirty years. In the 1970´s and 1980´s, industries like nuclear power plants and aviation understood the importance of considering socio-technical systems and the relevance for safety research (see Fahlbruch & Schöbel, 2011) and (see ICAO, 2009). The major catastrophes in the aerospace industry in this time changed the views on safety as well. While it appears that the Tenerife accident provoked a rethinking, specifically in the field of the human factor in the year 1977 (see Weick, 1990), the electrical power industry experienced in the year of 1979 the Three Mile Island accident with the same conclusions (see Perrow, 1994).

Whereas a reactive approach of "trial and error" or as in aviation known as "fix and fly" belongs to the no longer current methodologies - still the problem of assessing an uncertain future persists. Especially for aviation Roelen and Klompstra state: *"As of yet, the operational experience with measuring the effectiveness of safety management systems is very limited and there are many questions yet to be answered on measuring safety performance, demonstrating compliance with safety management regulations and the relation between safety management, quality management and safety culture."* (cf. Alfred Roelen & Klompstra, 2012: 1).

What once was a good start, namely accident investigation for safety aspects, now seems to be becoming a boomerang for aviation. The focus is still too much on incidents and accidents as extraordinary events (cf. Lofquist, 2010: 1522). Almost 30% of those interviewed in this study stated that their senior management does not demand analysis of safety performance indicators as a reporting method. The survey also confirms the aforementioned quote from Roelen and Klompstra which showed that while some proactive performance indicators get tracked like audits and regular reviews, yet the actual application shows a wide range. For example, the answers to the question of how often SPIs are updated varied from ad hoc (33%) to weekly, monthly bi-annually up to 33% updating SPIs annually.

From a thematic point of view, research initiatives in the helicopter sector focus mostly on the maritime and offshore market, supporting the oil and gas industry. Above all, two organizations were found contributing to aviation safety science. Those are the Netherlands Aerospace Centre (NLR) and HeliOffshore, which is a global safety focused association for the offshore helicopter industry. HeliOffshore is initiating and financing studies in the safety sector, also employing innovative methods from outside the aviation industry.

The target groups of this paper are HEMS organizations belonging to the legal category of "commercial air transport". Other sectors are general utility, law enforcement, offshore, corporate (VIP) and news gathering (see Seymore, 2018). There are also associations in the HEMS business, but no studies initiated by them could be publicly found. Although there are studies founded e.g. by the European HEMS association EHAC, those papers are generally not published. The availability of only few studies regarding the HEMS sector might also find its causation in the fact, that this sector only represents a small part of the total market. The HEMS regime in central Europe represents a cross section of patient safety from a medical point of view and aviation safety as well as a spectrum from sea rescue to alpine rescue work. Apparently the intersections of these different areas have barely been addressed yet. No paper could be found addressing SPIs specific in the HEMS domain, regardless of the fact, that the helicopter safety statistics show a 10 year high - as can be seen in the graphic below.



	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Serious Incident	2	0	1	1	3	1	2	3	3	7	8
Non Fatal Accident	8	3	6	2	5	3	4	8	3	4	9
Fatal Accident	2	2	2	2	1	0	1	2	3	1	2

Figure 5: accident statistics for helicopter commercial operations, other than offshore (EASA, 2020b)

Steffen Kasper et al. (2019) addressed a gap in the missing academic approach in aviation safety and presented a review of the existing aviation safety metrics. This was the foundation for a four-year project entitled: *"Measuring Safety in Aviation – Developing Metrics for Safety Management Systems".* The results of this project will be presented in 2020. In the initial paper they identified that the"*[…] long-established view on safety as absence of losses has limited the measurement of safety performance to indicators of adverse events (e.g., accident and incident rates)"* (cf. Kasper et al., 2019: 3)*.* They found, in a review of literature and industry references, six different definitions of safety and various approaches on determining the level of acceptable risk (cf. Kasper et al., 2019: 7).

In the aviation relevant literature no evidence could be found on methods to classify performance indicators or approaches to link measurement methods to the classifications.

Kasper et al. (ibid) found the following recent approaches in the field of aviation:

- Di Gravio, Manchini, Patriarca, & Constantino (2015) finding means to develop proactive safety indicators based on Monte Carlo simulations
- Karanikas (2015) coming from a psychological approach considering the controllability of events
- Bödecker (2013) using again frequencies and risk level from inputs in reporting systems
- ARMS (2010) using a combination of an event risk classification ERC (historical) and a safety issue risk assessment SIRA (proactive)
- The Aerospace Performance Factor APF used by EUROCONTROL using outcomes calculating risk over time

Apart from studies that were made in the field of aviation, there is a considerable amount of literature available in the general scope of safety research. Particularly probabilistic approaches were frequently found in the domain of process engineering. In those papers different concepts can be found like the “safety II” idea, which focuses on what went right versus the original understanding on analyzing what went wrong which is called safety I (see Hollnagel, 2014).

Apart from the scheduled airline business, there is consensus in these papers that SPIs are still not used sufficiently in the industry and that the performance indicator there needs to be operationalized more precisely (Kasper et al., 2019), (Lofquist, 2010), (Alfred Roelen & Klompstra, 2012),(Alfred Roelen, Lin, & Hale, 2011).

No evidence could be found in literature for an advantageous use of categories of SPIs, what is scrutinized in the first hypothesis and discussed in chapter 4.3. On the contrary: John Wreathall describes in a review about a paper called *“Thinking about Process Safety Indicators”* by Andrew Hopkins the downsides of getting too much into distinguishing between leading and lagging indicators. He takes an approach saying that data based on indicators is not responsible for improvements in safety. It is the safety management system behind it. This view is also partly represented by the system level model which is illustrated in Figure 10. In the model that was set up for this work, both the necessary background of an SMS and the advantages and necessities of a classification are shown.

The aforementioned paper of Kasper et al. (2019) was particularly interesting for this study. The initial paper and the four-year study carried out raises very similar problems only without relation to helicopter aviation.

In the helicopter regime, SPIs represent a tool that is used, but in fact, as the study has shown, the application is rather unstructured and at a very basic level. While the term “measuring indicator” initially leads to a thought of a reactive method, there could also be a preventive, even a predictive component to it, depending on what is measured. The second round of the survey showed that the median of percentages of proactive safety performance indicators is at 33% (Figure 6). Only two operations stand out with more proactive indicators than reactive which focus merely on tracking historic events without preventive character. In the system level model (Figure 10) reactive indicators are used to measure the validity of the proactive indicators which means that a balanced ration between those two categories of indicators seems reasonable.

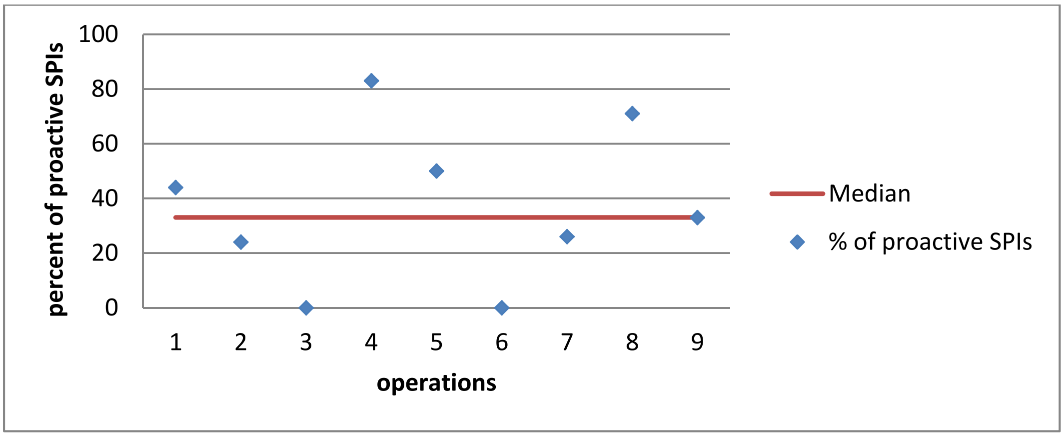


Figure 6: Distribution of percentages of proactive SPIs (own representation)

The two organizations with zero percent stated lack of time and personnel resources as reason for not using proactive SPIs.
The surveyed organizations primarily named audits and regular reviews as their preferred method to monitor safety within their operation. This confirms Lofquist's research, in which he argues that existing literature is no help in applying new methods of measuring safety as most of the literature is still focused on accident causation (cf. Lofquist, 2010: 1527).

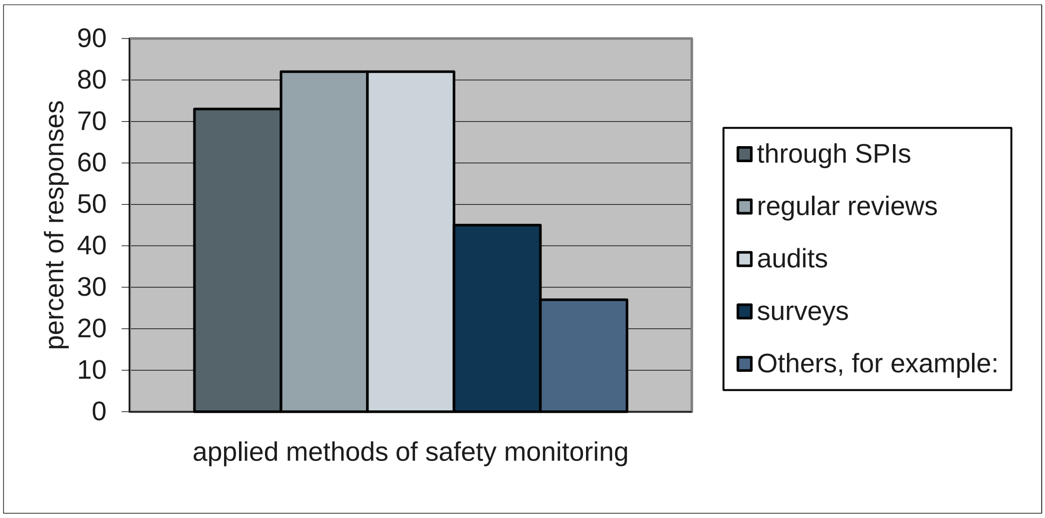


Figure 7: shows which methods have been specified for safety monitoring (own representation)

The importance of SPIs is mostly rated as "high" by the experts, but anyhow 73% stated SPIs as a method to monitor safety, from which can be concluded that the organizations with zero percent proactive SPIs in the diagram above (Figure 6) not only do not measure proactive SPIs, but probably no SPIs at all.
The delta between the 73% of organizations using SPIs for safety monitoring and only 64% using historical data could be due to a high amount of proactive SPIs but also due to a misconception of the expressions used, as one organization claimed not to use SPIs but actually uses numbers of past events in the company as a method for risk assessment which de facto is a kind of SPI. Two organizations stated to apply SPIs but do not use the number of past events as a method of risk assessments.

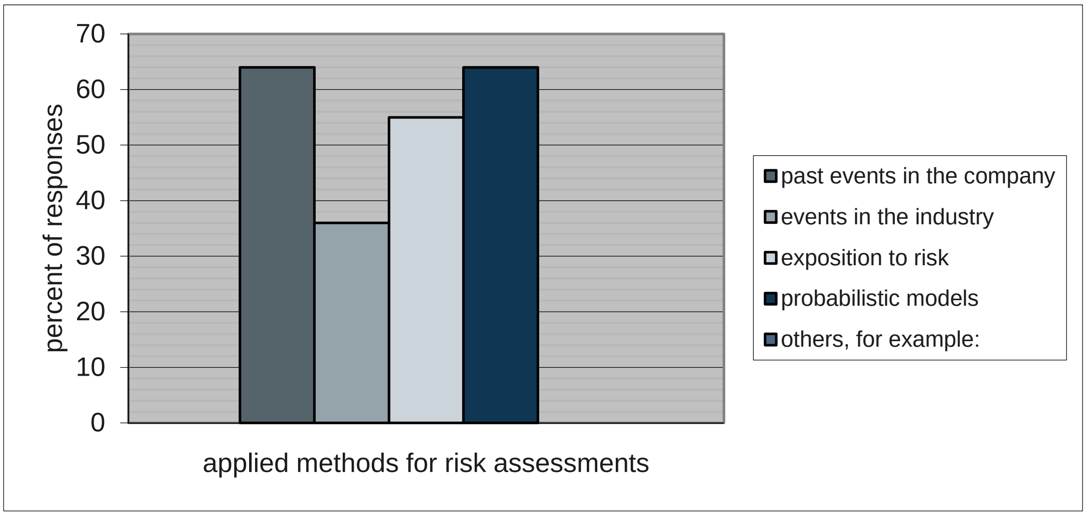


Figure 8: shows what the organizations evaluations on the severities are based on (own representation)

The European aviation authority EASA requires national authorities to adapt their auditing of operators into the direction of a risk based oversight. Accordingly audits and measurement planning are driven by a combination of risk profile and safety performance (see EASA, 2016). This puts the focus on the management of risk rather than fulfilling merely compliance obligations. Figure 8 shows what risk assessment in the HEMS business is like and what the organizations' evaluations on the severities are based on. John Adams titled his article in the Risk Management Journal with: “Risk management: It's not rocket science... It's much more” (see Adams, 2007). In general, the different approaches basically come all down to a cycle framework like the respective ISO standard for risk management which presumes the steps of risk identification, analyzation and an assessment of risks (cf. Austrian Standards, 2014: 6). In the available literature different approaches to evaluate risks and safety were found. Besides qualitative models which employ hazards analysis and hazard operability, there are quantitative models which use measuring indicators (cf. Adebiyi, Charles-Owaba, & Waheed, 2007: 178). Also the in aviation more common documents like the ICAO Doc 9859 “Safety Management Manual” recommends a similar cycle, first determining safety objectives and then using defined SPIs to measure identified actions (cf. ICAO, 2009: 4-2). The aforementioned document, as well as newer literature, recognizes the necessity of comparing leading indicators to the outcome measured by lagging indicators. Traditionally accident and incident rates have been the prime indicator for safety performance.

For example, a typical indicator in scheduled airline aviation is fulfilling the requirements of a stabilized approach. Therefore, certain criteria like an airspeed +/- 20kt of a reference airspeed, the correct landing configuration, a sink rate <1000 ft/min, the aircraft on the correct flight path with only small changes in heading and pitch usually have to be set, before descending below 1000ft above airport elevation (cf. Flight Safety Foundation, 2000: 2). For commercial operations, airports have to be pre-assessed, hence altitudes and headings are well known. Deviations from the exemplary aforementioned parameters can be tracked through flight data management (FDM). FDM means analyzing data from a flight data recorder which records the aircraft's flight data and depicts discrepancies from preset data like the parameters for a stabilized approach. Parameters like these are not applicable in the HEMS regime as an approach will - due to changing meteorological conditions and different obstacles - look different for every single landing.

In a typical aviation accident investigation historical parameters are the first which are checked to reconstruct the settings of the accident. Although an accident is always a deviation disturbing an existing safety equilibrium(cf. Alfred Roelen & Klompstra, 2012: 6), James Reason already doubted if the reactive search e.g. for the explanations for accidents was pushed too hard (Reason, 2008: 109 ed seq.). The focus on parameters which can be controlled and changed prevents to track the causal chain too far back in order not to get lost in details which will not help to increase safety. Also Eric Lofquist (2010) questioned sarcastically the art of measuring nothing in a changing civil aviation industry. In his paper he describes the paradox of measuring safety in an ultra-safe industry as an outcome measure using traditional measuring methods. Guidance Material for commercial aviation regulations define acceptable safety targets for HEMS missions as less than 5×10^{-8} (cf. EC, Regulation 965/2012: GM1 SPA.HEMS.100(a)). Six years earlier, René Amalberti defines ultra-safe as the risk of an accident to be 10^{-6} or lower (cf. Amalberti, 2006: 253). According to his contribution to the book "Resilience Engineering" he questions the general classification of aviation as ultra-safe. In his analysis he attests scheduled passengers operations a risk magnitude below 10^{-6} and helicopters, scheduled flights and business aviation a magnitude below 10^{-4} (ibid). While an important point is made in the incomparability of the different aviation sections and the fact that there are areas of aviation which have a different exposition, he fails on the comparison of figures as number of departures and flight times are mixed up. This gives also an impression of the difficulties to get dependable data in aviation.

Reliable numbers and statistics for the helicopter emergency medical service in Europe, which could be used for a risk based approach by the operations, do not exist. EASA is publishing an annual safety report, but unfortunately even EASA is not very persistent with its numbers. There are barely numbers for HEMS and if a closer look is taken, it can be found that the 2006 EASA annual safety report counts six fatal accidents with helicopters and the 2007 report only lists 4 of those accidents for 2006 (see European Aviation Safety Agency, 2007). The same thing happened in the year 2011. Consequently those numbers are not reliable enough for statistics or for risk based approaches. Following own estimates, more than two thirds of the commercial helicopter companies in Europe can be classified as small or medium size enterprises (SME) with less than 250 persons employed and less or equal than 50 million euro turnover (classification according to Eurostat). Also the survey shows that more than two thirds of the organizations are below these numbers and have nine or less aircrafts. More than half of the operations complete 5.000 or less flight missions a year compared to more than 25.000 missions in the bigger operations. The size of the operations matters in terms of collecting safety data. Having a considerable portion of SMEs with presumably few data sets on occurrences, a database for risk based approaches needs to have publicly available valid data from similar operations. This needs a common approach.

4.2. Safety Management using Safety Performance Indicators

The use of SPIs does not make sense for every operational question. But with adequate performance indicators, different causes of uncertainty can be checked in a frequent iterative process, in addition to the legal obligation to use SPIs already mentioned above. 73% of the organizations taking part in the survey for this study stated that they are using SPIs already. However, much of the data used is historical data that does not contribute to a proactive approach. In Figure 1 only the approaches mentioned in the survey are shown, without taking into consideration that most of the methods can also be interpreted in SPIs. Other methods mentioned were: Reporting systems, investigations, inspections and risk assessments. The following chapter considers arguments why and how safety performance indicators can be applied.

4.2.1. Why utilizing safety performance indicators

There are mainly three reasons found in literature for using safety performance indicators:

1. To monitor the level of safety in an organization which requires valid data showing trends
2. To allocate resources and to decide where actions need to be taken; here, links between indicators and outcome are necessary
3. For reporting and to motivate personnel in management positions to take proper action (cf. Hale, 2009: 479)

In order to understand how existing mechanisms function and to be able to predict possible events, an understanding of the current system through the use of SPIs is required (cf. Herrera, Hollnagel, & Håbrekke, 2010: 1). Safety as meaning of a condition of certainty for an individual or system to be highest possibly protected from unjustifiable exposure to adverse conditions also implies that in order to be safe first the risks and possible exposures need to be considered. Safety is simply said the absence of risk while risk refers to the probability of harm occurring. This means that safety and risk counteract - safety increases, while risk decreases (see Figure 9). In helicopter aviation, especially in the field of helicopter emergency medical services, situations, where no complete set of information is available, are common. Having no

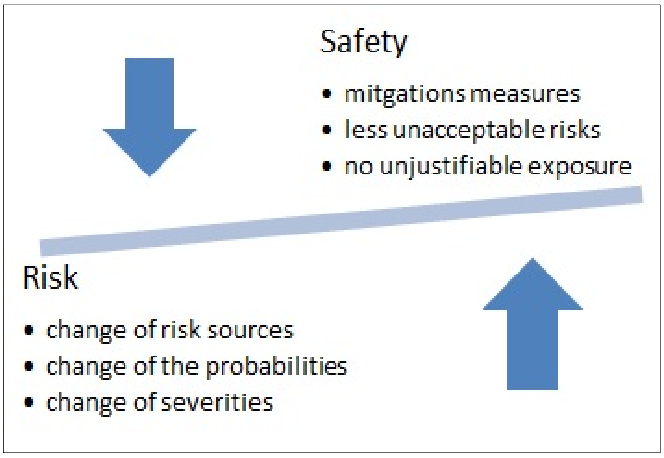


Figure 9: Correlation between risk and safety (own representation)

chance to pre-assess landing sites, not knowing where the next leg of the flight will go or having to be prepared for all kinds of traffic or weather situations leaves a high degree of uncertainty. The standard procedures and this degree of uncertainties define the risk profile which needs to be addressed in the risk management.

Peter Drucker, pioneer in modern management, said: “If you can´t measure it, you can´t manage it.” (cf. Drucker, 2009: 1). However, measuring alone carries the risk that outputs will be accepted as facts without questioning their validity. This would mean using only the lowest subsystem in the system level model (Figure 10) without reaching the levels of interpretation an application of the outcomes. Therefore, SPIs need to be tied to safety objectives or to safety performance targets. As safety is about allocating the right resources to the right operational area, a review should also be done constantly in order to check whether parameters are measuring what actually is searched for. This also includes the elimination of wrong or corrupt data. For instance, what happens if flight data of helicopters is used and the data linkage to the systems stops working? It could produce false-positive results as no exceeding of flight parameters is detected. Another example of an external influence for possible false-positive results is a reduction in flight activities like during the corona pandemic in 2020 which showed a decline in rescue missions due to the lock down but a reduction in risk exposure can only be derived from the reduced likelihood. The exposure to risk per flight will not be reduced if not even raised due to less routine or other arising influencing factors like flying with protective clothes. The number of operations and incidents only correlate linear with very large numbers or reports and especially the numbers of usable incident are usually not available in HEMS operations.

The data acquisition and control is part of the lowest level of application in the aforementioned system level model, which is described in the next chapter as a subsystem of the parameters. The data gained is necessary for the implementation of higher system levels in which the measurement of SPIs is used again as the basis for risk management and for interpretation. Is a low risk level measured by lagging indicators due to good management or is it due to external influence factors e.g. caused by a reduction in flight operations? For this interpretation reference values and leading indicators need to be integrated in the overlying subsystem of performance indicators (Figure 10). Risk management utilizing safety performance indicators is a method of using mostly numerical values to assess where resources should be prioritized and to establish a systematic reporting. At best, SPIs are reproducing a trend in key attributes of the specific system and create an accurate picture of the exposure and the culture of handling risks within the organization. Therefore, SPIs can be regarded as means to monitor and also as a tool for effective communication of safety parameters, e.g. management reports or external and internal communications. For helicopter emergency medical services SPIs cannot be duplicated from other areas of aviation and are, due to the multiplicity of the set-ups, difficult to define. Depending on the phase of the mission, ranges for parameters need to be defined larger but still need to be able to depict situations away from clean model settings.

The utilization of SPIs tied to safety objectives also provides the possibility to define escalation factors. For example a certain degree of exceedance could be managed by routine procedures while major exceedances result in an obligatory report to the senior management or constrains and limitations on actual operations. A system of SPIs can be flexibly adjusted and shows outcomes and changes quickly (cf. Janicak, 2015: 26).

4.2.2. System and context boundaries

Catastrophic man-made disasters in the aerospace industry, like the crash of the space shuttle Columbia, the collision of two Boing 747s at the Tenerife air accident, or the Überlingen mid-air collision always changed the considerations to the possible source of error and showed very clearly that the aerospace regime depicts a complex sociotechnical system. ICAO found a trend from the technical era until the late 1960s, to the human factors era, considering the man/machine interface until the early 1990s which advanced across the organizational era considering organizational culture to the prevalent era of the total system approach (cf. ICAO, 2009: 2-1 et seq.). The definition of complexity in the specific system of SPIs is found in the various subsystems and the sets of connected agents influence each other in a dynamic way - what can be found in figure 7 (cf. Haberfellner, de Weck, Fricke, & Vössner, 2012: 51 et seqq.).

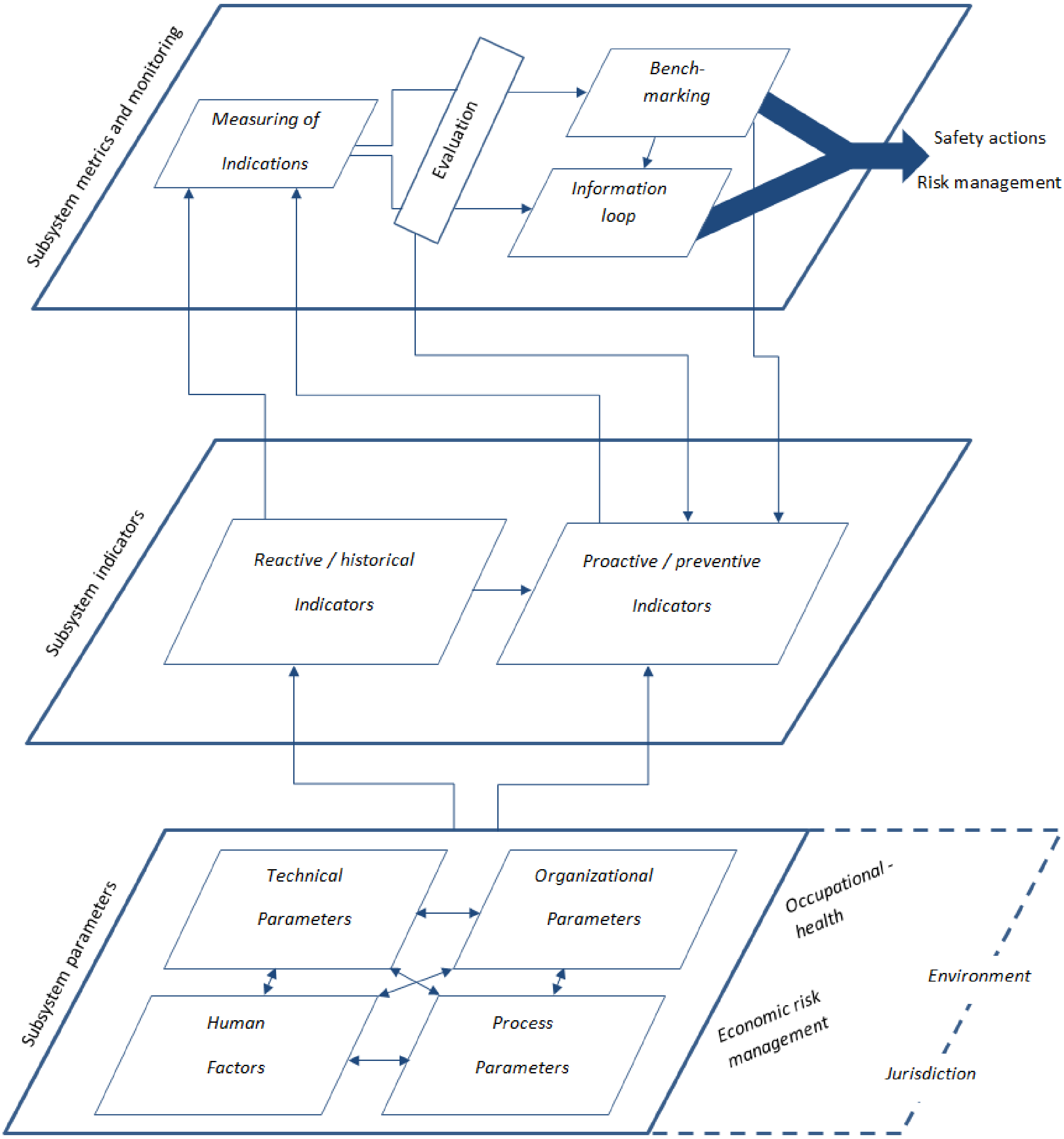


Figure 10: System level model of safety indicator application (own representation)

The multiscale nature and the fact that the process and its complexity are influenced by human and sophisticated technology shows the sociotechnical aspect of the system (see Alessandro, 2011).

Employing a system of safety performance indicators implies also a preparation of the subsystem of parameters and the subsystem of metrics and monitoring. Figure 10 shows the different subsystems where the foundation is made by the parameters which represent the data flowing into the system.

The lower rhomboid embodies the sphere of action, whereas the solid line sets the limits for the safety management's sphere and the focus of this paper. The dashed line shows items which are however inside the control limit of an organization but are usually outside the sphere of safety management. Factors like the in the figure

exemplary pictured jurisdiction or environmental aspects are outside the actual sphere of action of an organization but could hypothetically be indirectly influenced. For example, lobbying could influence politicians and justice or environmental or ethical factors could be influenced by cultural changes triggered through the organization. The dashed area is not included in the views in this paper.
As safety is an interconnected matter, of course all those parameters which are left out in this thesis, due to the excessive volume which would be necessary to describe all aspects, need to be considered to get a holistic picture of the safety and risks in an operation. For instance, safety has always been reflected in the context of economic success. Besides the business-management parameters, also occupational health and the in the HEMS industry very important medical aspects are also left aside in this paper.
Operative safety in this paper means all activities of a safety management directly connected with the flight, rescue procedures and maintenance operations. Accordingly, the field “technical parameter” in Figure 10 labels parameters in conjunction with equipment used or parameters describing component failures. Organizational parameters overlap in many areas, with human factors described as subject - there can be risk appetite and exposures to risk or requirements of productivity. It is a matter of safety culture if the goals and objectives of an organization are congruent with the goals and objectives of an individual. Therefore, the human factors parameters not only describe mental, cognitive behavioral and social factors of influence (cf. Badke-Schaub, Hofinger, & Lauche, 2012: 3-20), but also the accordance and commitment to organizational goals. The field process parameters subsume all indications which picture procedural matters of safety. This could be an overview of audits or its output or just the plain process of recognizing external factors like the registration of bird activities in the vicinity of landing sites. Another important parameter falling under the process type is the proactive monitoring of risk controls. Especially the small and mid-sized organizations lacking larger data sets to draw consumptions (cf. Alfred Roelen et al., 2018: 176-177) need to define processes for the monitoring e.g. through gaining external data. The monitoring of the effectivity of risk controls represents a possible newer approach to take into account, especially in absence of robust data bases. For a higher level of detail, a look into the sources of data like reporting systems or flight data monitoring systems can be made.
The middle subsystem in Figure 10 represents the actual managing and the classification of safety performance indicators. It is the content of the first hypothesis of this thesis which is described by this subsystem. The categories of SPIs play a central role in monitoring and measuring safety. Plain numbers are not effective if not compared with goals or benchmarks as well as the effectiveness of preventive indicators need proof of historical indicators to justify their application. This subsystem gets its data mainly from the parameters level of the system but should also get feedback information from evaluations and comparisons drawn. In this subsystem, the overlays and variations caused by individual SPIs are also detected. The use of categories makes it possible to compare SPIs on different topics in order to detect a build-up that could potentially push a system to its limits.

The top subsystem describes the processing of the gathered information and the outcome. This is the level where the application takes place. The measuring shows how parameters can be matched and conditioned for an evaluation. In a dynamic system a self-reflected evaluation will show that many approaches in illustrating indicators are inadvisable, as no significant prediction of safety can be found there, or the effort to quantify the parameter is disproportionally high. There is a gap between the application of mitigation measures and the safety objective, unless the level of the residual risk is below the level of the defined safety objective. This is the gap to find in the evaluation under the subsystem of metrics and monitoring, which means finding the effectivity of risk controls and finding out which safety measures are expedient. While most of the organizations in the survey state that they update their SPIs at least monthly, this model can be set up to deliver data in real time. The application on the subsystem of metrics and monitoring includes a comparison with the organization's internal goals and also a comparison with industry standards. Specific values for a systematic comparison could be published by authorities or independent institutes or a comparison of indicator levels could be discussed in associations emerging from the particular industry segment. In the area of helicopter aviation notable organizations are the Helicopter Association International (HAI), the European Helicopter Association (EHA) or specific for the medical service in a regional context stands the European HEMS & Air Ambulance Committee (EHAC).
This comparison with the total system beyond the organizations perception is also recommended by ICAO. As Figure 11 shows how from the beginning of the 21st century the development in safety monitoring advanced from a focus on individual safety performance to a wider context comparing with the total aviation system. (cf. ICAO, 2009: 2-2). The survey which was carried out as part of this study discloses, that in the helicopter field most actions are still made in the context of the human factors era and the organizational era. Bourrier already argues in her paper of 1998 that too often the organizational aspects are only analyzed after catastrophic accidents, while the proactive essential understanding of the "normal" operation is missed out (cf. Bourrier, 1998: 133). Although no data on helicopter aviation was found, the study

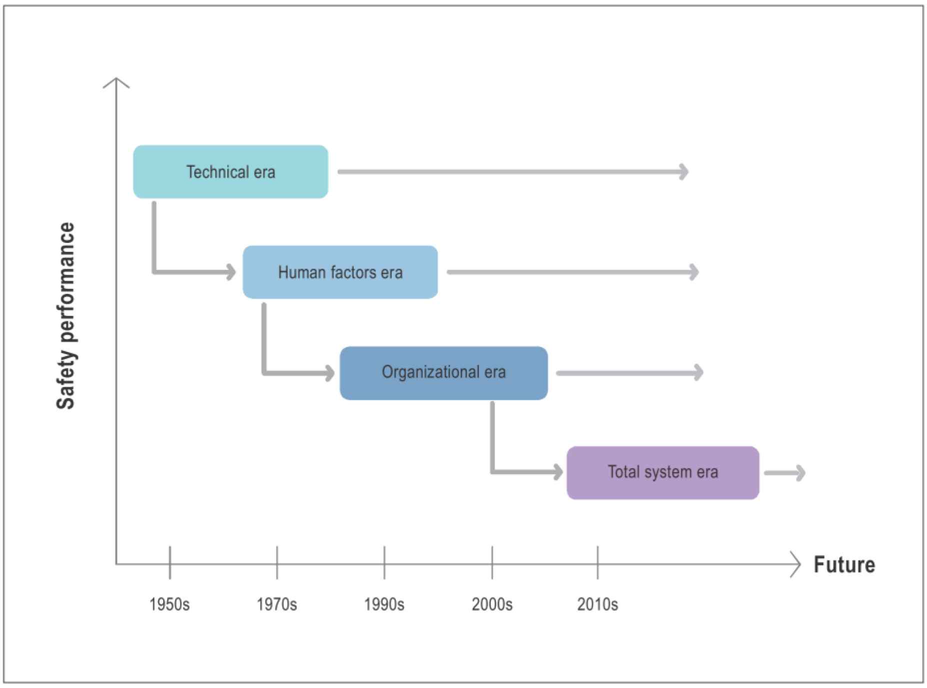


Figure 11: The evolution of safety as seen by ICAO in Doc. 9859 (2009)

showed that 80% of the HEMS operators surveyed track influences from the technical and human factors areas. With only 40% of companies tracking organizational influences, this sector of the industry seems to be on the threshold to the organizational era. After all, one organization stated that connections in the "total system" - thus failures due to the complexity of the system e.g. interfaces of organizations or influence of governmental administration are investigated, but contrary to that there was also an organization that does not yet consider any of these relationships.
The system limits are also defined by a legal framework which has already been mentioned in the definitions section. The decisive authority in Europe is EASA, which laid out the rules for safety monitoring under the EU OPS regulation ORO.GEN.200 (EC, Regulation 965/2012). In these regulations can be seen, that the approaches for safety and compliance monitoring are to a vast extend interdisciplinary. Particularly items like continuous improvement or management of change are a cross-sectional matter. In the beginning of the EU-OPS, EASA did not explicitly differ between SMS and QMS (cf. Alfred Roelen & Klompstra, 2012: 3). QMS stands for Quality Management System, which is nowadays in aviation referred to as Compliance Monitoring. In the regulations, which are in effect by 2020, compliance and safety are both described under management systems. Even though Kasper et al. said 2019 *“It is probably easier to find non-quality compared to unsafe situations…”* still compliance and safety management systems are based on the same principles. A number of authorities recommend systems, especially in the safety management as a certain degree of alignment with their mandatory state safety program could be achieved. For instance, Transport Canada as national aviation authority has a “Civil Aviation Integrated Management System Standard” (see Transport Canada, 2016). Integrated management systems coordinate and manage diverse activities in a coherent system. The use of integrated systems is recommended for companies in aviation and due to similar approaches a common use of standards in the safety and compliance area is obvious. In the survey organizations were asked about their management system. 82% of all companies claim to have integrated management systems. Interestingly also over 80% describe their compliance and safety as two complete different divisions. This could be due to divergent definitions of integrated management, but most likely it can be justified by the fact that, although large companies use integrated management systems, the departments remain completely independent, due to the size of the departments. Although also excluded in this paper it is important to understand the correlation and interaction of the different management systems as there might be overlaps between factors in different management systems e.g. occupational safety and compliance or aviation safety (cf. Alfred Roelen & Klompstra, 2012: 1-2).
Without showing the details of the subsystems, Figure 10 shows already a large number of elements interacting on different levels. Considering the detailed aspects especially in the parameters subsystem links can be dynamic and follow non-linear connections. This is particularly true for the relationships between indicators that measure certain parameters in the field of social-technical interaction. The description of the overall system depends on the level of detail in which the system is viewed. The in Figure 10 shown system is in detail a massively interacted, complicated system due to its amount of interacting elements and their

connectedness. But according to the system engineering model described by Haberfellner et al. (2012) the overall system cannot be considered as a complex. Therefore, it would require a dynamic between the subsystems in the system which does not exist. This is important, as stable procedures in the overall process are necessary to measure safety performance. Defined safety performance indicators (e.g. for resilience) can contribute to this necessary stability although more than half of the companies surveyed stated that there are areas in operations which cannot be measured by SPIs.

4.2.3. Requirements for useful indicators

The underlying question is how likely something is to happen and what the consequences of the event are. Severity and probability are traditionally the basic parameters of risk management. In advanced applications factors like broken barriers or exposure rates play a role as well as how risk is perceived or what could be the “unknown unknowns”. The items of uncertainties for a specific operation, for instance the not pre-assessed landing site at night, could be how high the probability of colliding with powerlines is or how adverse cloud conditions could be detected if night vision imaging systems fail. The “unknown unknowns” are those areas where no knowledge about the theoretical presence of a risk is given. Cengiz Turkoglu, senior lecturer at the Cranfield Aviation University explicated at a speech about aviation safety management in Vienna in 2018 that, in recent years, two contradictory safety theories have been observed in the aviation industry:

- The first theory is that there are no new accidents and basically the same root causes are leading to the same kind of accidents.
- The second theory describes that with sufficient safety management systems all future accidents will be based on unforeseeable circumstances – so called black swan events.

On the assumption that in a functioning safety management system where uncertainties are tracked and control barriers are based on the findings, an evolution of the prediction model can be found. This prediction model including the aforementioned theories would have to be dependent on the efficiency of the SMS. On an accident rate up to 10^{-3} the next accident will be a repetition of the previous accident, on a rate of 10^{-3} to up to 10^{-5} the next accident will be a recombination of part of already existing accidents or incidents, in particular using the same precursors and at a rate of 10^{-6} or better, the next accident has never been seen before. The latter represents the aforementioned black swan events. Its decomposition may invoke a series of already seen micro incidents, although most have been deemed inconsequential for safety. (cf. Amalberti, 2006: 270) Safety performance indicators represent the assessment of the known uncertainties. For instance, as mentioned above, helicopter operations know that there is a higher risk to perform night landings to unassessed landing sites. Known uncertainties would for example be represented by the fact that meteorological phenomena can’t easily be seen at night or that the

airborne assessment of a landing site during night entails risks that are not a factor during daylight assessments.
The unknown uncertainties are represented by black swan events. Incidents which have not happened before and which have not been foreseeable. Unfortunately for this study, it was not possible to get accurate data on accident statistics in terms of accident investigations where the root cause factors of incidents or accidents were new and therefore would denote black swan events. Neither EASA nor large helicopter manufacturers could provide figures for this study. Manufacturers claimed that those numbers are collected but are confidential due to the fear of legal consequences. According to the survey, 78% of the organizations have at least one to two black swan events per year. Taking a look at the helicopter accidents of the last decade published online by the Austrian Federal Investigation Authority, a variety of repetitive suspected causes can been found (BMK, 2020) which therefore cannot be classified as black swan events. From the initial need to measure security, there arises the need to find indicators (cf. Øien et al., 2011: 150). These indicators need to be suitable for comparing existing system elements, but can also be helpful in creating situations that prevent black swan events, especially in the field of safety culture. Following Amalberti's line of arguments, the aviation industry is considered as an ultra-safe industry and should, according to an incident rate of less than 10^{-6}, mainly fall in the category of new events where the root cause has not been seen before.
But just as the entire industry cannot be compared, individual operations cannot be thrown together. A typical, even by EASA often used, specification to compare is the sole figure of an accident rate for a specific operation type which might be an interesting number for an insurance company or a striking value for possible customers, but it does not offer any usable content based on which a management could focus and allocate resources on specific items. Most HEMS companies in Europe do have a plurality of work operations e.g. standard HEMS flights flying from a mission site to a hospital as well as special maneuvers like night vision operations or human external cargo or hoisting operations or flight training for the recurrent training of their flight crews. Flight training for instance has a complete different risk assessment and requires a complete different set of SPIs. For a functioning risk management system requirements must be met, that measuring systems can measure SPIs, both in individual categories and the combination of latent conditions. Almost every catastrophic failure in high reliable organizations is based on a chain of errors which represents a combination of broken barriers (cf. Reason, 2008: 11 ed seq.). This search for multiple causes is probably the most difficult challenge of SPIs to define. The systems describing multiple cause models are called epidemiological models such as Reason's well known Swiss cheese model (cf. Reason, 2008: 12). Compared to singe root cause models, epidemiological models use categories like active failures through actions and inactions and latent conditions (cf. Kasper et al., 2019: 11). Next to single cause models and epidemiological models, are the systemic models on which a closer look is taken in chapter 4.
Looking for characteristics of useful safety performance indicators very general popular management criteria can be found like the SMART principle standing for specific, measurable, achievable, relevant, time-bound (cf. Janicak, 2015: 49 ed

seq.). While of course these definitions always make sense, Thomas Rockwell (1959) went into more detail. He described the characteristics of good measures of safety performance as following:

- Quantifiable and permitting statistical inferential procedures
- Valid or representative to what is to be measured
- Provide minimum variability when measuring the same conditions
- Sensitive to change in environmental, or behavioral conditions
- Cost of obtaining and using measures is consistent with the benefits
- Comprehended by those in charge with the responsibility of using them (cf. Alfred Roelen & Klompstra, 2012: 7)

Herrera et al. (2010: 5) found in a study additionally to that the characteristics of ownership, which supposes that there is personnel which is measured, also owns the respective indicators in terms of responsibility towards the outcome. Another characteristic that is not mentioned in Rockwell's enumeration is reliability, defined as*: "The data should as far as possible be either objective or intersubjectively (sic!) verifiable."* (ibid).
Based on the model in Figure 10, the initial steps when creating a SPI based performance measuring system is to deal with the subsystem of parameters. In the literature also notes on the amount of SPIs can be found. SPIs always have to be tied to a safety objective as well as to the available data and the reliability of measurements and then the number results automatically. The amount should allow statistical analysis of changes and trends, but too many SPIs might distort the view and result in a lack of focus (cf. Alfred Roelen & Klompstra, 2012: 7) and (Kjellén, 2009: 487).
Having a defined safety target also means that an acceptance criterion exists. Measuring uncertainties, the acceptance criterion will most likely be a function of a probability of occurrence. In the *f-N* curve it is depicted as frequency dependent on the extent of significances (cf. Bai & Jin, 2016: 716). There have been attempts to set common risk levels outside of aviation for instance in the Netherlands a maximum of acceptable individual fatality risk of 1 x 10^{-6} was suggested (cf. Ale, 2005: 240) for the population on a state level. But in aviation the loss of an aircraft could result in complete different risk acceptances for a big operator with several hundred aircrafts or the risk acceptance compared to a small operator where thBe loss of a single aircraft could mean bankruptcy. Also, an a-priori assessment of a risk could change after a series of incidents which would also require an operation to adapt or set SPIs in place determined for the circumstances of these incidents. Therefore, there is no such thing as a common risk level in aviation: *"The level of*

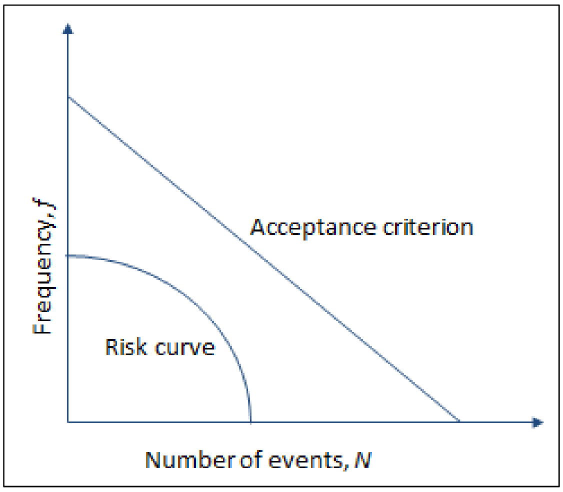


Figure 12: *f-N* curve (cf. Bai and Jin 2016: 716)

accepted operational risk has not been universally established." (cf. Kasper et al., 2019: 6). States in their role of oversight and organizations have to define their own risk acceptance which again makes it hard to compare (cf. ICAO, 2009: 9-17). Judgements on the frequency of occurrences rely again on the availability of data. This might be limited by the possibilities and constraints of the industry. 82% of the surveyed organizations for this study claimed that an exchange of operations' statistical data like mean time to failure or incident rates would help their safety management. Limiting factors in the HEMS regime could be the diversity of operations. Although the basic concept remains the same in the following phases: flying to a mission site, patient treatment, transport to an adequate hospital and finally the flight back to the base, the details in the phases can differ greatly including completely different risk assessments. While a HEMS flight in the Alps holds a high probability of having to rescue a patient from an inaccessible area using human external cargo flights (HEC), a HEMS flight in the lowlands could include flight with no visual reference solely by instruments (IFR) or extended flights over water. This also goes hand in hand with various mitigation measures, such as certain kind of external cargo equipment or autopilot systems. Therefore, useful indicators ideally divide in a set of indicators which are comparable on a cross company basis to be able to exchange performance information supporting risk assessments and a second set of indicators which are precisely adapted to the specific requirements of the organization.

Based on a solid structure of the lowermost subsystem (Figure 10), parameters are created which deliver data into the overlying system of the different categories of indicators.

4.3. Categories of safety performance indicators

To find ways to define SPIs, it is first of all important to distinguish between the different classifications, as categories contain different goals that can be pursued with it and also offer different reporting options and different opportunities for cross-company exchange. Also, when answering the question which SPIs reflect a real image of the safety situation in an HEMS operation, categories also play an important role. Compared to the system model in Figure 10, this means that SPIs at the parameter level are intended to illuminate individual aspects. The merging of SPIs for the analysis of a specific situation in a defined sub-area of the operation takes place one level above, on the subsystem of the indicators. This approach can also be found in the classification into "high level indicator HLI" and "low level indicator LLI". According to a framework for the nuclear industry, HLI are used by management for decision making and LLI are on a technical detail level and therefore assigned to a specific person in the process (cf. Øien et al., 2011: 154). Categorizations into HLI and LLI can be helpful for system understanding, but no advantages in terms of an exchange or for monitoring methods can be seen.

While the classification of SPIs itself was already questioned (see Wreathall, 2009) as to possess the risk of over-administration, there are reams of evidences in literature on how to use certain SPIs and how to take advantage of their measurements. Although it is not explicitly mentioned, these studies always assume a previous classification in the application of indicators. An important use of categories that is mostly found in the literature is the comparison of leading and lagging indicators. Lagging indicators provide information of historical events, and they are therefore evidence of whether measures have taken effect. Hence it is the initial comparison of leading and lagging indicators which gives an initial understand whether the right measurements have been taken. This method requires that safety actions are measured by proactive leading indicators as a form of active monitoring (cf. Øien et al., 2011: 158). The terms of leading and lagging can be found in different variations as it can be seen in Table 2.

Although in the literature lagging indicators are often depicted as outdated or even negative, they cannot be neglected. The typical lagging indicators are ever so important as they serve as the necessary reference value with which the taken measurements can be compared. While those reference values in aviation are typically related to flight time e.g. “number of accidents per million flight hours” in the HEMS business a relation to flight operations e.g. “number of accidents and incidents per take-off and landing” or “per flight leg” seems appropriate due to the short flight times per mission. Depending on the parameters also a calendric reference might be reasonable e.g. “number of occurrence per year”.

leading indicator	lagging indicator
upstream	downstream
predictive	historical
heading	trailing
positive	negative
active	reactive
predictive	retrospective
input	output
driving / Monitoring	lagging
proactive	reactive
activity	outcome

Table 2: pairs of indicators (KASPER et al.2019: 8)

Another important way of basic categorization is the distinction between quantitative and qualitative indicators. In this study the focus is on the group of quantitative indicators, because it is this group that can be subjected to the mathematical, analytical measures and provides an easy application for comparison. 73% of all companies surveyed use only SPIs with numerical values and more than half of the experts questioned state that there are areas which are not measureable with SPIs. Under the items mentioned in the survey as example for those areas not to be measureable, were medical aspects or safety feedback from crews as well as findings from line checks. Line checks are annual check flights *“…to ensure the crew member´s ability to operate efficiently under normal conditions, whereas other checks and emergency and safety equipment training are primarily intended to prepare the crew members for abnormal/emergency procedures.*” (cf. EC, Regulation 965/2012: GM1 ORO.FC.230). When selecting SPIs it needs to be decided if the evidence must be generalizable form or if a descriptive image would describe the requirements best (cf. ICAO, 2009: 4-5). ICAO also highlights the advantages of a combined approach between qualitative and quantitative safety performance

indicators and warns of problems with a single approach as not all situations might be depicted (ibid).
All of the above points are closely related with human factors and safety culture and it seems that it is difficult for companies in this area to find measurable SPIs. The discrepancy with no unique standards for safety culture were the basis for the research by (see Montijn & De Jong, 2006) who developed the Aviation Safety Culture Inquiry Tool (ASC-IT). This tool supports organizations in the assessment of their safety culture and provides six categories for safety culture characteristics with the accompanying, measurable indicators. Independently from any tool used, categories to depict organizational soft variables like safety culture or culture changes within socio-technical organizations are paramount. The importance of such categories could also be shown in the investigation of the space shuttle accidents (cf. Lofquist, 2010: 1522).

4.3.1. Applications of SPI categories

A single safety performance indicator cannot be evaluated if it is only an isolated figure. Only the connection with safety targets and the integration into the monitoring system create the necessary reference and a usability of the value as depicted in Figure 1. Although there has not yet been an evaluation of classifications of SPIs in literature, the following items were noticed during the research for this paper, which makes the application of classification advantageous:

Utilization of categories in a monitoring and management system
In comparison with the system level model (Figure 10), indicators classifications have to be found as a prerequisite for the design of a safety management system. Different technical parameters for instance might be measured using the same tools e.g. maintenance actions using tables or enterprise resource programs (ERP). This monitoring of technical parameters can be established rather quickly and immediately offers the prerequisites to be compared across companies. Already a comparison at the parameters can be beneficial. For instance the fault rates of certain components can serve as an SPI. The survey showed that 73% of the organizations are already tracking the time between component failures and all organizations claimed to be willing to share those kinds of operations statistical data (there is no clear connection between the size of the company and the willingness to exchange information).
The system level model (Figure 10) also shows the need for classification. Compared to the commonly used flight data systems (FDM) in aviation, no use of categories would mean that hundreds of data sets from a flight are collected. Each considered individually contains no meaningful information about the prevailing situation. Every indicator must be assigned to a safety goal, and the interpretation of achieving a specific safety goal will require several SPIs combined as a group. Assuming that the measurement of SPIs primarily serves the application of safety barriers and the recognition where these barriers have to be used, the question also arises at which level of the system level model safety barriers are set in place. For technical and process parameters barriers can be found in the subsystem of the parameters

whereas the decision to implement barriers e.g. for found interferences in the training area involve the monitoring at the subsystem of metrics and monitoring and the interpretation of historical and preventive SPIs. This requires knowing which measurement values can be assigned to the training area, the associated group of people and the aircraft group used and which culture parameters.

Use of categories in reporting

In order to obtain resources for certain safety measures, purposeful management information is required. Eight out of eleven organizations in the survey senior management demand SPIs as a reporting method. This is evidence that there is an overall knowledge about SPIs and their importance to objective safety monitoring in the current management positions of HEMS companies. In these operations categories of SPIs have to be applied for reporting in addition to subsets of the most critical indicators (cf. OECD, 2014: 29) as most of the safety activities are directed in general terms of safety activities and not on a specific measured SPI. Especially for managers who are probably not familiar with the details of the subsystems and the parameters behind them, a formation of categories is necessary. For instance crew communication training can be a safety activity which results out of several indicators in the field of human factors, showing a noteworthy trend that communication has been a critical item.

An example for reporting using SPIs can be found in Figure 13. In this illustration, the risk indicators of incidents in an operation are risk values derived from a likelihood-severity matrix. Those values are summed up and presented as an "overall risk indicator". The peaks in the curve are due to the fact that the risk matrix used has an exponential structure. As mentioned before, linking with safety goals is paramount and is done here by setting maximum levels and escalations limits. In this example those alert levels were defined by the single and double standard deviation which ensures the dependency on the overall distribution. This simplistic representation of several different complex interdependent events requires little previous knowledge and the general risk situation can be quickly understood. According to ICAO (2009: 4-15) all events that go beyond the arithmetic mean should be investigated. Not only the amplitude of single events but also the variability of the chart gives an impression of the safety state of the system.

The second important task of reporting does not concern reports to the management but the information to employees in order to create a safety culture. In both applications adequate analysis of the gained data need to be done. While knowledge about safety performance monitoring can be taught in trainings, the display of safety performance in order to create awareness for safety needs to be easily understandable without pre-knowledge. 55% of the surveyed organizations already use SPIs for safety information, while the remaining 45% have a variety of safety information but without using SPIs. Questions about the knowledge and the training on the interpretation of SPIs were not part of the survey.

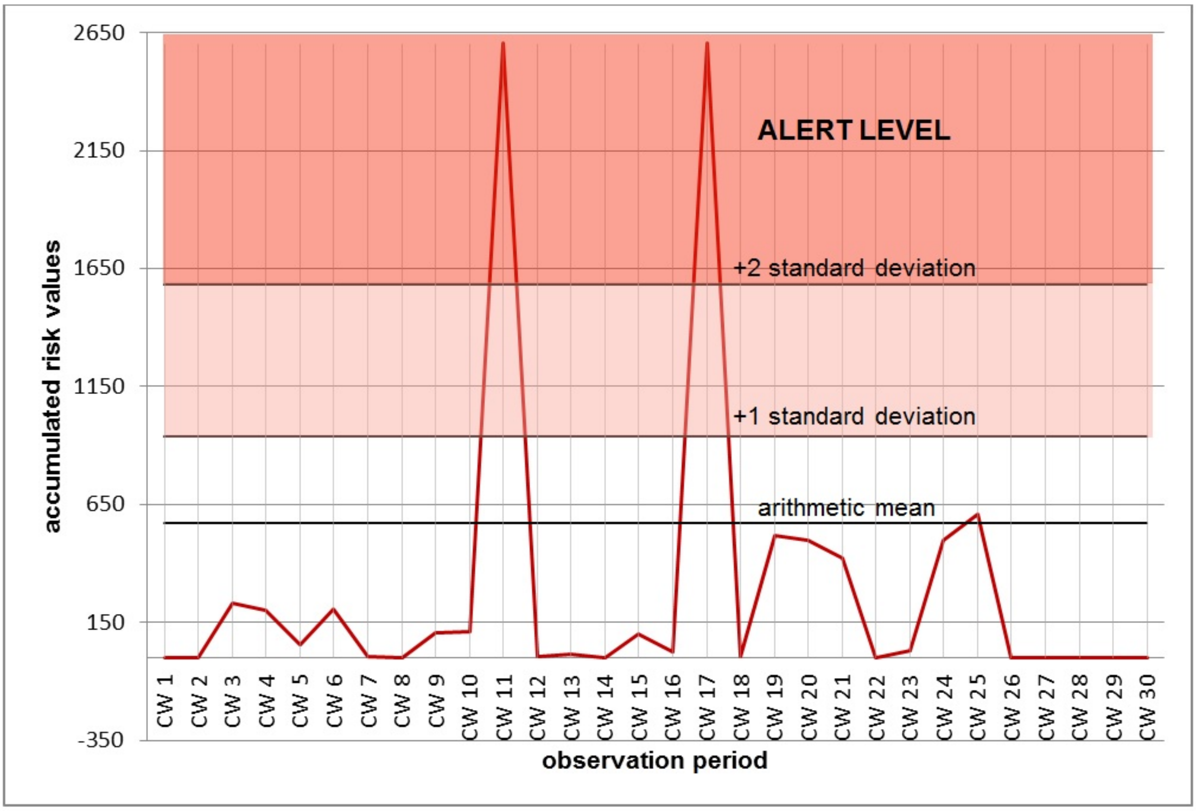


Figure 13: Example of a representation of risk levels using SPIs (own representation)

So there are two different applications of safety information. On one hand, for management to be able to recognize where resources need to be allocated and, on the other hand, for employees to be able to see where particular threats are and how important the own contribution to the security of an organization is.

Comparison of categories on a cross-company level

The comparison of performance indicators or specific values of categories of SPIs can help with the evaluation of exposures to risk. The median number of aircrafts in the HEMS operations surveyed is nine and only three of them have more than 20 helicopters. This means, average HEMS operations are too small to get usable numbers for a valid assessment of risks. The comparison of SPIs corresponds to customary benchmarking in business, with the challenges of comparing complex relationships that arise in the areas of various management systems and air rescue procedures. Especially numerical values derived from technical applications can be used for benchmarking, as they are most easily comparable, but then the approach is subject to the usual disadvantages of benchmarking, which is a high complexity compared to other analysis methods and above all it has no predictive character (cf. Holloway, 1995: 148). To search for accident precursors and use the predictive value of cross-company exchanges, information of leading indicators needs to be shared. It is the knowledge of black swan events, which of course cannot be measured by SPIs at the time of occurrence. In order to exclude the disadvantages of classic benchmarking, classifications need to be determined on the smallest common denominator and set as an industry wide standard. All surveyed organizations would be willing to share data, 82% of the experts interviewed stated that the exchange of operational statistical data would help their safety management. While the exchange of data on the level of technical parameters can be done easily, it is much more difficult in those areas that contain soft variables like cultural aspects and therefore depend heavily on the methods used. Companies in the same industry might be holding back to exchange detailed data since they are most likely in a competitive situation as well. It requires a high level of understanding from senior management to commit to an exchange of data and recognize the benefits. A categorization also makes sense here, since agreement on certain category parameters allows a degree of anonymity and prevents high effort for otherwise quite complex benchmarking. Benefits of this kind of benchmarking for example can be: values such as incidents on landing sites per 1000 missions which can result in an adoption of standard operating procedures (SOPs) as well as in altered methods of influencing factors regarding the selection of landing sites.

A different approach is necessary in the categorization of indicators in the area of human factors or in the field of change management. These types are usually in their structure strongly dependent on the configuration of the management system which again can be very different from company to company. The indicators found will show little congruence and will therefore not represent a valid comparison. The circumstance that it is difficult to find comparable numerical values is also reflected in the survey in which a third of the respondents stated that they measure change management, but do not use SPIs for this type of safety action, let alone a cross-company comparison. Alfred Roelen et al. (2018: 177) found in a research of state of the art academic literature, aviation industry practices and documentation published by regulatory and international aviation bodies that*: “ […] there is a lack of standardization across the aviation industry regarding the development of safety metrics […]”*. Although the study of Roelen et al. focuses on the airline industry the reasoning also applies in the HEMS sector. Neither ICAO nor EASA have a definition for the expression "risk control".

Even if indicators seem to be alike, different interpretations and measuring methods can make a comparison impractical (Alfred Roelen & Klompstra, 2012: 8).

Although 82% do not use external indicators for comparison, the survey showed that there is a demand for exchange of data, especially for operations statistical data and safety management contribution and productivity as well as an exchange of risk exposure data. In the survey, the organizations in particular rated the authorities at European and national levels as trustworthy sources. Due to the market density, this would definitely be a task that would fall into the responsibility of EASA. Unfortunately, previous attempts by EASA have been heavily influenced by the typical official procedures. Accidents in the sense of the ICAO definition were used as data basis and classifications were classified according to the incident reporting catalog ECCAIRS. This again narrowed the possibilities down to reactive methods. (see Network of Analysts, 2014)

Although, according to the survey, the confidence in the associations (60%) was a quarter less compared to the authorities, at least the specialist knowledge of the HEMS internal processes and the possibilities to create proactive indicators overall appear to be greater.

4.3.2. How to find appropriate categories

The question is which risk factors and also success factors should be queried in the company's own monitoring system, in order to get the desired picture of the safety situation in the respective operation. Opposed to the definition of individual indicators which will be discussed in chapter 4.4, the definition of the categories primarily serves the application of the aforementioned areas such as reporting, the cross-company comparison and the installation of a monitoring system.

The formation of categories corresponds to the subsystem of the indicators in which the synthesis takes place. It is then to be found in the individual setting which classification suits a particular operation best. At this level an adoption of specific needs has to take place, as every operation will have different requirements in order to be able to use the resulting monitoring for its safety management system. As an initial approach to find categories influencing each other, a risk map like the one shown in Figure 14 can be used. The boxes in the graphic represent some of the main themes of a HEMS operation; those of the same color belong to a certain risk owner who is responsible. For instance the risks in the area of flight operations (green) fall under the responsibility of the flight operations manager as risk owner and the economic risks (blue) in the area of management. The size of the boxes shows their interdependence and therefore also gives an impression on their vulnerabilities in case on failing of an element. For instance the box depicting a safe flight operation is connected to almost any other major system element and therefore it is the biggest element whereas only few other elements are directly influenced by the system's element political stability.

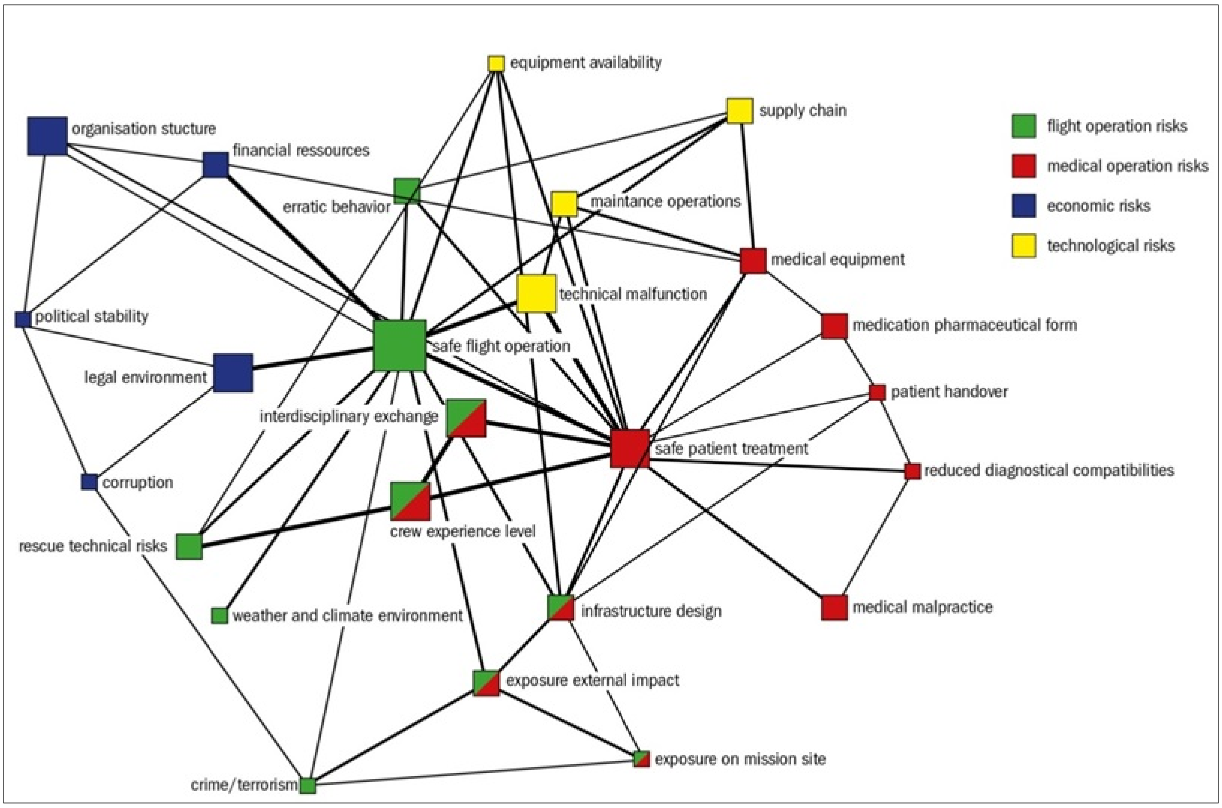


Figure 14: example of a risk map for a HEMS flight operation (own representation)

Recognizing these interdependencies is necessary in order to get through clustering from the subsystem of the parameters into the subset of the indicators. The risk map shown here is at the level of an entire organization. In a further, more detailed application of this system (e.g. the system element “crew experience level” is itself analyzed by a working group for its dependencies) categories are created that can be used for interpretation. The “crew experience level” might then deliver categories like “recurrent crew training” and “regular crew checking on normal and abnormal procedures”. For extensive use, the SPIs in these categories are both reactive and proactive. Using the "Crew experience level" as an example, the reactive indicators would measure how often a lack of experience of the crews could be found as the root cause of incidents. Potential scenarios identified as possible black swan events or experiences of other organizations in the industry could serve as source for preventive indicators. The found elements and links of this method can be grouped to clusters which will be measureable by the same means. This method in a more detailed level can also show the overlapping, non-linear connections and dependencies.

A more detailed view of categories delivers categories which will be adequate for a cross-company exchange on the lowest applicable level. The comparison on an SPI-level will not work in most cases, because a direct comparison is due to the variety of non-standardized measurement methods and different management systems most likely counterproductive.

The aforementioned more detailed use of the risk map in Figure 14 could also be utilized in probabilistic approaches. Even though 64% of the surveyed organizations state that probabilistic models are already used in their risk assessments, these approaches are merely being represented by the judgement of likelihood based on case numbers (see Kurpiers, 2018). Although event sequence diagrams (ESD) like fault trees and influence diagrams have their fixed place in aviation accident investigation, the actual use of probabilistic models in the sense of probabilistic risk assessments (PRA) or Bayesian Believe Networks (BBNs) are usually not used in HEMS. The federal aviation administration (FAA) of the United States developed the hybrid causal logic (HCl) as a method which uses the latter additional means beyond the conventional risk analysis techniques to define a framework in which certain risks can be viewed as hazard (cf. Alfred Roelen et al., 2011: 7). However, this model was not made for operational use, but for the FAA principal inspectors to exercise safety oversight. Due to the complexity of such models, corresponding resources are probably only available in larger HEMS operations which can provide the corresponding human resources in the safety department.

In a further step, a harmonization of categories across the HEMS industry can be considered in order to exchange data with other operations. As mentioned in the chapter on comparison of categories, a close look must be taken at which classifications are suitable for comparison. Some risk influencing factors might not be directly comparable or directly measureable. A definition for a RIF is: *“A risk influencing factor (RIF) is defined as ‘an aspect (event/condition) of a system or an activity that affects the risk level of this system/activity’ ”* (Øien, 2001: 7). It might need several different SPIs as an operational variable to measure such an RIF (ibid). It is therefore not only difficult to differentiate what is actually measured, but also a high level of detail of information that, when published, can possibly give a competitive advantage to other organizations. This can be a reason why a possible comparison at the level of categories could possibly be too imprecise and the scope for interpretation too great to be able to make effective use of it. In order to minimize the level of detail and still keep the benefits high, a different categorization can be considered for the cross-company comparisons. Apart from the values for risk assessment, the focus here is primarily on preventive factors and the creation of resilience. Accordingly, a classification in terms of the risk management cycle or according to its impact would also be conceivable as depicted in Table 3. An exchange e.g. in the area of hazard identification could also contribute to create equal competitive conditions as the same standards are used.

Table 3: alternative options for classifications

classification according to the occurrence	classification according to risk management cycle	classification according to impact
Operation´s statistic data	Hazard identification	safety objectives
External factors	Risk Management	risk level
safety culture	Safety Performance	sustainability
human resource	Safety Improvement	insurance costs and ratings
exposure to risk		resilience and preparedness
safety management		competitiveness and image
resilience indicators		staff motivation

4.4. Defining of safety performance indicators

Compared to the previous chapter in which categories and the risk factors were in the focus, this chapter deals with the underlying indicators, thus the lower subsystem in the system level model. The fundamental question in the definition of SPIs must be which goal the measurement is pursuing. It will be a matter of the strategy of an organization to decide what the safety goals are and based on these goals, the detailed safety targets and the initial associated safety performance indicators including their categories can be derived. To structurally search for defined indicators different methods are available. Priority was given to methods by means of which non-linear relationships can also be represented, since the overlapping of processes is an important circumstance that must be recognized by a system utilizing SPIs.
In the following, some of the methods are analyzed which can be used to delineate SPIs. The methods of functional resonance analysis method (FRAM) and risk influence factors modeling (RIF) represent very sophisticated approaches that mainly make sense for detailed processing or for gaining methodological knowledge. Since many operations do not have sufficient resources available, simpler forms will be

discussed in the second part of this chapter on how existing practices of analysis can be used to find valid SPIs without having the benefit of large amounts of safety related data.

4.4.1. The functional resonance analysis method (FRAM)

FRAM is a method for modelling complex socio-technical systems. Based on the understanding that mistakes and success are based on the same causes and that accidents are always complex in nature, Hollnagel (2012) developed the functional resonance analysis method. In the case of incidents usually, a linear system is used where it is assumed that complex problems can be broken down into individual elements. Those bimodal elements resolve in true or false or work and fail and the probability of error can be described individually. FRAM assumes that in everyday performance variabilities arise resulting in resonances which are non-linear combinations of unexpected elements. (see Hollnagel, 2012)
It is the aim to dampen these resonances in a system by considering the characteristics of the dynamics and dependencies of the current system. It consists of 4 steps which are intended to dampen the variability that daily performance is subject to:

- Characterization of the purpose of the model: it can be used in a traditional view to describe accidents or as a source for finding latent conditions, finding of important system functions as well as describing them by means of six basic characteristics (called aspects: time, control, output, resources, preconditions, input)
- Description of the potential variability of the functions, understanding how functions are interconnected and how this can lead to unexpected results
- Characterizing functional resonance, the system can describe what happens if an element or component fails
- Finding of ways to monitor and identifying effective countermeasures or barriers (cf. Hollnagel, 2012: 36)

The aspects are usually drawn in a diagram as can be seen in Figure 15.

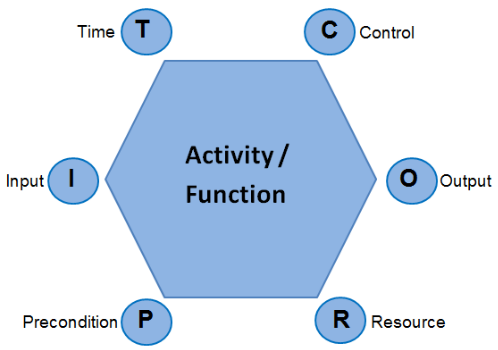


Figure 15: function FRAM diagram (Hollnagel, 2012: 46)

While the FRAM analysis represents a good option for the identification of leading indicators this stage still needs the information of historical data as every function has to be assessed to its potential variability. After considering the variability of the individual aspects, it is about recognizing the resonances between the functions. This reveals how the variability of a function affects other functions or the entire system and an overlay of variability of different functions can also be recognized. Like in the reporting tool mentioned in chapter 4.3.1 a high level of resonances can give information about a safety system that reaches its limits. Recognizing the variabilities gives the chance to set SPIs first in order to measure functional resonances and secondly to put measures into effect, which in turn are then monitored by indicators. (cf. Herrera et al., 2010: 6)

Figure 16 shows an example of an FRAM analysis, which was made to analyze a mid-air collision trying to apprehend the relationships between human interaction, technical equipment and air traffic control. This example shows very illustrative the correlation of a socio-technical environment and the characteristics on resilience of air traffic control (see De Carvalho, 2011).

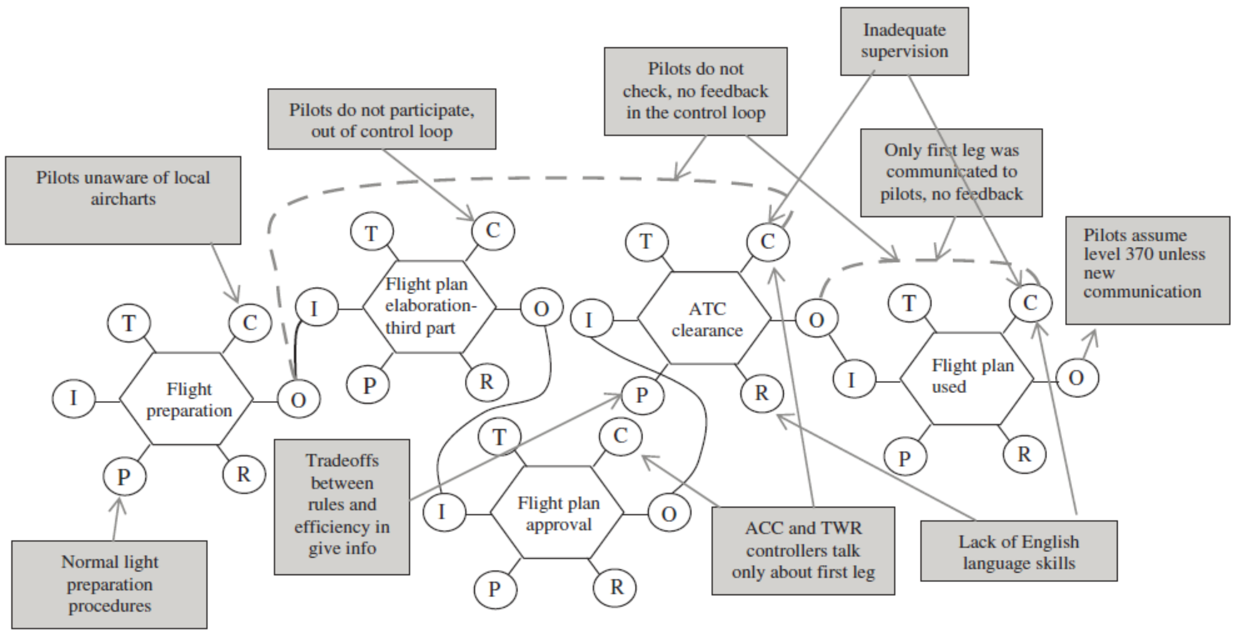


Figure 16: diagram describing a take-off (De Carvalho, 2011: 1490)

As this example shows, FRAM analyzes can be used in retrospective and prospective and, accordingly, can also be used to identify leading and lagging indicators. Only a few methods are suitable for finding leading indicators and so an increased operational use can be beneficial for HEMS operations. Apart from one, none of the organizations in the study knew about the method of FRAM analysis. In the literature research only FRAM analysis on an academic level were found. A single study with operational impact was found in which a FRAM analysis of helicopter flight operations offshore on the Norwegian continental shelf was performed. The author of the FRAM model was also acting as a co-author of this study.

4.4.2. Risk Influence Modeling (RIF)

The risk influence model is a tool to map a total risk picture and the risk influencing factors (RIF) in order to get an overall picture of the risk at all levels and to identify the most effective risk reducing measures. "*A risk influencing factor (RIF) is defined as 'an aspect (event/condition) of a system or an activity that affects the risk level of this system or activity" (Øien, 2001: 63).* The necessary risk influencing factors build a grid out of relatively stable factors affecting the risk of an activity (e.g. transport of a medical passenger). Starting from a specific accident category, those factors' frequency and accident consequence have to be identified. Relations between a main objective and the RIFs of the model are used to predict effects of possible changes within the system. (cf. Hokstad, Jersin, & Sten, 2001: 313 ed seqq.)
Although the model contains predictive options and a possibility to include probability calculations of certain risk influence factors, it is still based on valid data material on past events which needs to be available. While all levels of an operation can be depicted in the diagram it is limited in the amount of details. This creates the necessity to use more detailed methods for critical risk influence factors.
Figure 17 shows an example of an RIF diagram which was made to present a general picture and the influencing factors for the north sea off-shore helicopter transports. After defining a category, the RIFs were divided into three levels according to their direct impact. (Hokstad et al., 2001: 311)
The resulting cause-effect relation suggests the influencing factors that must be monitored using SPIs.

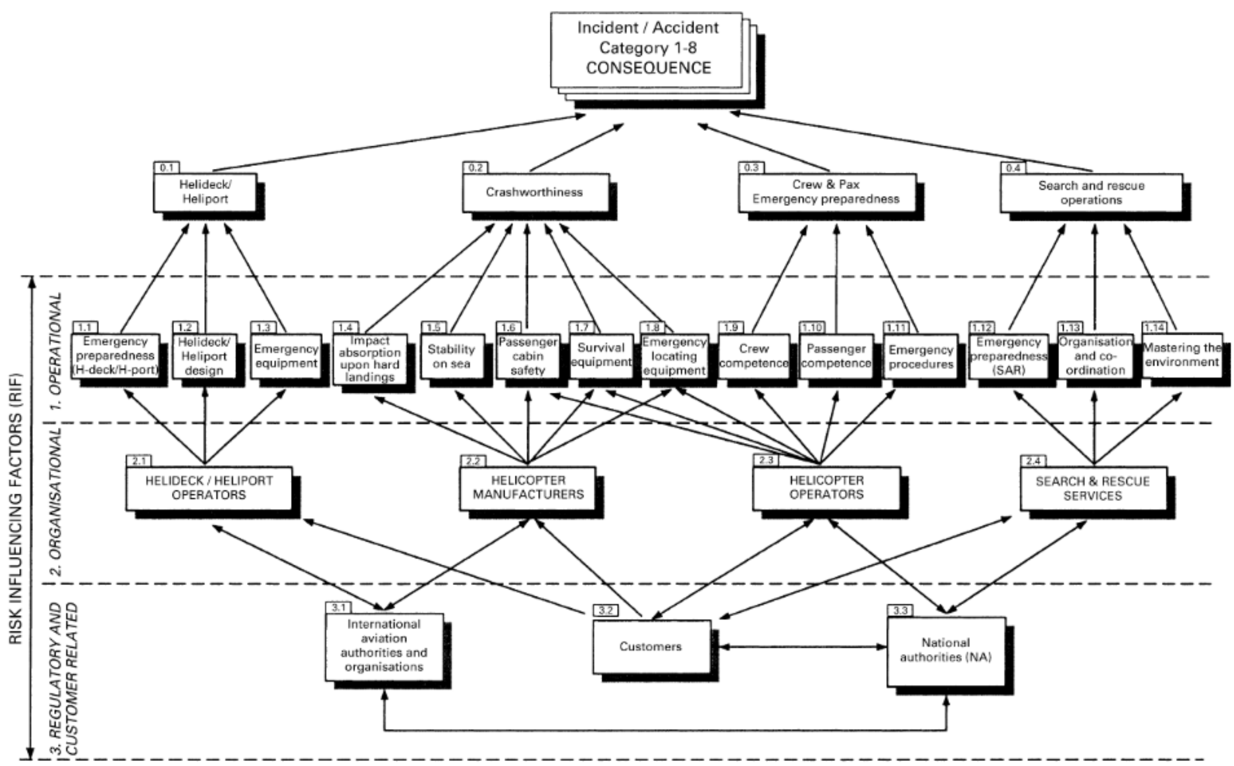


Figure 17: Influence diagram according to (Hokstad et al., 2001: 314)

4.4.3. Simple forms to systematically define SPIs

The different approaches in an HEMS operation will always make a combination of methods beneficial. As mentioned, it will often be due to the availability of data sets on the basis of which the methods are validated and thus always include a reactive component. It lies in the application of the methods to find proactive performance indicators and ultimately to be able to also measure resilience independently from the frequency of the occurrences.

Resilience could also mean preparation of measuring methods which spring into action after an event to mitigate the consequences but are set in place before an event.

Methods that are already in use in HEMS operations are the bow tie analysis and the Swiss cheese model for the investigation of accidents.

The bow tie analysis is an established graphic method to assess risk situations which is mostly used preventative. In the center of the analysis is the event to be examined, to which, in a logical link, the threats and consequences of the occurrence of the event are added. The control barriers added on the threat side are intended to minimize the probability of the occurrence to happen and the ones on the consequence side are the control barriers which come into force after an event in order to reduce the consequences. The barrier measures not only serve as elements that are monitored by the means of SPIs, but can also be clustered according to the categories of the risk cycle into: avoided, reduced, transfer and residual risk. The measures on the consequence side will most likely fall into the regime of resilience management.

A similar concept of safety barriers is described by James Reason (2008: 10) in his book “Managing the risks of organizational accidents”. The model is widely known as Swiss cheese model as it is best described as several slices of cheese as symbol for defenses which have to be arranged in a way, that one series of holes is congruent. This would allow hazards or a so called accident trajectory to pass through all barriers. While the first defenses are labeled as active failures, which are made by the personnel, the later ones are mostly classified as organizational errors. The Chernobyl RBMK accident is a tragic example where defenses were deliberately removed in order to execute tests. This window of opportunities that a set of holes allows a hazard to develop represents a so called latent condition. (cf. Reason, 2008: 10 ed seqq.)

Finding the latent conditions is a major part of the accident investigation but it can also be applied to smaller occurrences and thus find out about the necessary barriers which need be observed or, in a more dynamic view, which barriers are not allowed to be arranged in a way that they give way to an accident trajectory. SPIs can then be defined by the set-up of the barriers as defenses. It could be the barriers itself which can be quantified by SPIs or it could be several critical control points which define the barrier. The synopsis about the dynamic of barriers would then be part of elements in the subsystem of indicators as well as elements in the subsystem of metrics and monitoring.

The Swiss cheese model also serves as foundation for the so-called “SOL – safety through organizational learning” method which was developed for the nuclear power industry and provides a scientific event causation model (cf. Fahlbruch & Schöbel, 2011: 28).
Both approaches, the bow tie analysis and the Swiss cheese model, help to contribute to an overall system and help to understand isolated elements. But both methods are inefficient to depict complex safety issues (cf. Lališ, 2017: 51).
Another possible way of generating parameters without big data sets is the usage of SPIs based on the effectivity of risk controls. Alfred Roelen et al. (2018) are describing what the definition of the effectiveness of a risk control can look like. Basically safety barriers like mentioned above in the bow tie analysis are measured in terms of how often they break as a risk control and how often the barrier was successful.(cf. Alfred Roelen et al., 2018: 181)

4.5. Impact of safety performance indicators on operational safety

Particularly variations in human performance have always made it difficult to measure safety performance in socio-technical environments. Although the literature gives no clear indication of whether the presentation of SPIs creates an advantage in a safety culture, it can be assumed that a generally understandable disclosure of information will have a major impact on a safety culture. However, several publications point to the use of safety data to support organizational learning and pro-active monitoring (Fahlbruch & Schöbel, 2011) and (Gerbec, 2013).
Gerbec (2013) proposes in his paper an approach to support organizational learning by comparing measurements and outcomes in a safety management system. This approach largely depends on the assessing of safety performance indicators. The comparison of activities observations and outcome observations is based on five steps which are also depicted in Figure 18:

1. Implementation of an incident-investigation procedure at company level
2. Monitoring of the safety preventive activities performance observations using various tools (in order to pro-actively reveal deficiencies)
3. Identification of outcome deficiencies observations using the procedure (step 1), including root-cause analysis for the process-safety incidents/accidents
4. Comparison of activities deficiencies (step 2) with outcome deficiencies (step 3) in order to reveal latent deficiencies and to prove the relationships
5. New universal safety indicator method is proposed to enhance organizational learning from both activities and outcomes observations comparisons (step 4). (Gerbec, 2013: 1114)

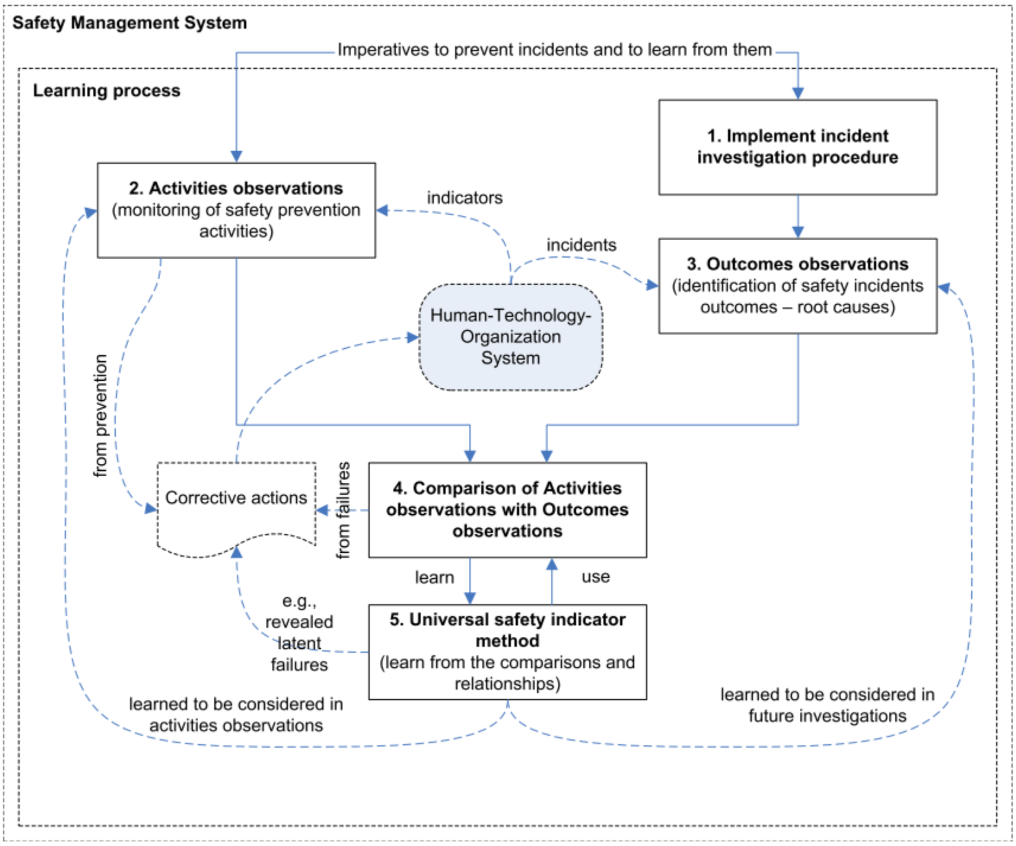


Figure 18: organizational learning through observations of activities and outcome (Gerbec, 2013: 1114)

On one hand safety as a supposed soft factor is often viewed as very subjective, on the other hand mere listing of incidents, as is usually done by authorities (cf. EASA, 2020b), is not legitimate for assessing realistic security performance.
According to the system level model (Figure 10), SPIs provide a detailed view of the safety situation on the level to be considered and permit statistical calculations like the creation of a safety performance index similar to the presentation method that was presented in the chapter 4.3.1.
One safety performance index using SPIs is the so called Aviation Performance Factor (APF) which helps getting an integrated view on safety data and allows deeper analysis of safety performance(cf. Lališ, 2017: 52):

$$AFP = \frac{\sum_{i=1}^{k} Wi*Ni}{appropriate\ denominator} \qquad (1)$$

As already described in chapter 4.3 denominators like “missions flown” or “number of flight legs” seem to be appropriate in the HEMS regime rather than time intervals. The “W” is the weight of the respective SPI which first has to be weighted by experts, “N” depicts the number of indicator observations and “k” represents the total number of SPIs in the system. (cf. Lališ, 2017: 52)
An application of a six-sigma capability index, which has so far mainly been used in the process industry, could also be considered for aviation. It measures how much a process moves away from the desired normal value and how much it is under control. *“The Capability Index tells you if the process is truly capable of meeting the outcome requirements.” (cf. Janicak, 2015: 39)*

The key of SPIs for operational safety lies in finding the SMS elements which can eventually provide an indication of an accident. In HROs single events can deliver a clue for a latent condition (cf. Øien et al., 2011: 159). Therefore event analysis and statistical calculations use safety performance indicators through setting barriers which are again monitored by SPIs or through re-design of a system and its monitoring.

The probably biggest impact of SPIs on operational safety is the knowledge that can be gained from the effectiveness of the risk controls. This finding will give a profound knowledge about where risk controls have to be set and how close to near accident the causal chain went before a barrier stopped the accident trajectory. Thoroughly investigated it can also provide information about which kind of barriers are the most effective and which kind of failures e.g. organizational or process failures represent the currently greatest threat for the operations.

5. Empirical approach

After the findings of the literature study, the approaches of the HEMS companies in Europe were researched in an iterative process. The decision to select a Delphi analysis was due to the possibility to identify and qualify expert-opinions to the topic of safety performance indicators. It was not the aim to create consensus but as the application of SPIs is a rather less used method the technique to collect the answers and give a summary back to the expert group had the opportunity to find new insights and come up with new ideas.

The knowledge about the application of SPIs in the HEMS regime gained can be above all found in the context and the correlation of the answers given in the survey. However, special care is required here, since only a few organizations work in this environment and the number of samples is not sufficient to make substantiated statistical statements. Nevertheless, there are connections that allow conclusions to be drawn even from small numbers of cases. The correlations stated below are results of the analysis after the first round of the survey which were summarized and given back to the respondents to be challenged in the second round. The results after the completion of both rounds are also included in these subsequent items.

The general legal basis that applies to HEMS organizations stipulates that there are safety and compliance departments that are described in the management systems (cf. EC, Regulation 965/2012: ORO.GEN.200). A trend, integrating those two functional units into one management system comes initially from the field of air traffic management. Not only does it stimulate an efficient use of resources, it also helps in the application of performance indicators as very similar requirements to performance indicators underlie both departments. Applying an integrated approach only gives a statement about the procedures used. It does not necessarily mean that compliance and safety are one department. The smaller companies also claimed that the application of SPIs is a matter of resources that are not always available.

Following the survey two out of eleven of the questioned organizations apply an integrated management system. The same amount of respondents claimed that safety and compliance are two separated departments. Organizations where compliance and safety are not two separated divisions are in the segment of the smaller companies of the survey with less than 10 aircrafts which is most likely related to the effort involved. What is striking is the fact that it is the largest organizations that do not use integrated management systems which may be because of a disproportional greater effort that is incurred by large companies when the management system is changed. Another commonality is that in this comparison the larger organizations are the proactive applications of SPIs through regular reviews. It is the smaller companies with no separate divisions for safety and compliance that also use these proactive methods. One reason could be that those smaller companies have integrated management systems but also quite often the functions of the safety and compliance manager are combined in one person. Thus the in quality and compliance management popular method of reviews is simply transferred to the safety part. The bias in this conclusion is that the larger organizations in the survey include corporate companies as well as governmental organizations. While the legal basis is the same when flying HEMS, regardless to whether it is done by a governmental organization or not, the approach to management systems done by governmental systems might be different.

Table 4: Correlation of organizations sizes with integrated management approaches

How many helicopters are currently active in your operation?	How many flight missions does the operation you work for achieve on average during one year?	How many employees are currently in the operation you work for?	Do you use an integrated management system?	Are compliance and safety two separate divisions?
v_1	v_8	v_9	v_17	v_18
3	1.000 - 5.000	8	Yes	Yes
3	1.000 - 5.000	75	Yes	Yes
9	1.000 - 5.000	126	Yes	No
20	15.001 - 20.000	370	Yes	Yes
90	> 25.000	900	No	Yes
ca. 50	> 25.000	ca. 1200	No	Yes
17	10.001 - 15.000	250	Yes	Yes
9	1.000 - 5.000	150	Yes	No
2	< 1.000	40	Yes	Yes
2	< 1.000	30	Yes	Yes
54	> 25.000	250	Yes	Yes

The work done like surveys, analysis and audits naturally also depends on the available human resources. Almost half of all organizations surveyed have one full-time equivalent workplace for safety. Although a third of the respondents fly more than 25,000 missions a year, there is no organization with more than 3 full-time equivalents.

An apparently important fact is the correlation of the organization's size with the application of safety monitoring. But again due to the small number of samples no correlation could be found between the parameters of the size of an organization and the application of monitoring safety. All organizations surveyed stated that they are monitoring safety but the methods are distributed non-linear through the use of SPIs, regular reviews, audits, surveys - or 27% admitted to methods like investigations, inspections and reports which are predominantly reactive methods. Most organizations in the survey conduct between two to five audits per month which shows that audit seems to be accepted as a preventive tool. The applications of audits do not correlate with the size of the organizations or management approaches - which could be assumed as audits are originally a tool that has emerged from quality management.

The search for specific safety performance indicators also showed no significant trend in the samples. Exemplarity were the historical indicators of mean time between failures (which includes all kind of failures e.g. organizational failures or human factors) and mean time between component failures (this only includes technical failures) used for the survey. Most of the surveyed companies do track those very common indicators, two companies do not track them and another two do keep track of either one of those indicators.

The application of SPI categories seems to be common in the HEMS aviation. The statistic is led by the internal statistical data which probably represents the category with the least effort for the measurement (see Figure **19**). However the requirements of the senior management in an organization seem to produce side effects. Two out of three organizations where senior management does not require SPIs do also not apply categorizations. Also the applications of SPIs which consider an accumulation of events, which are customary reporting factors, are not applied in those organizations.

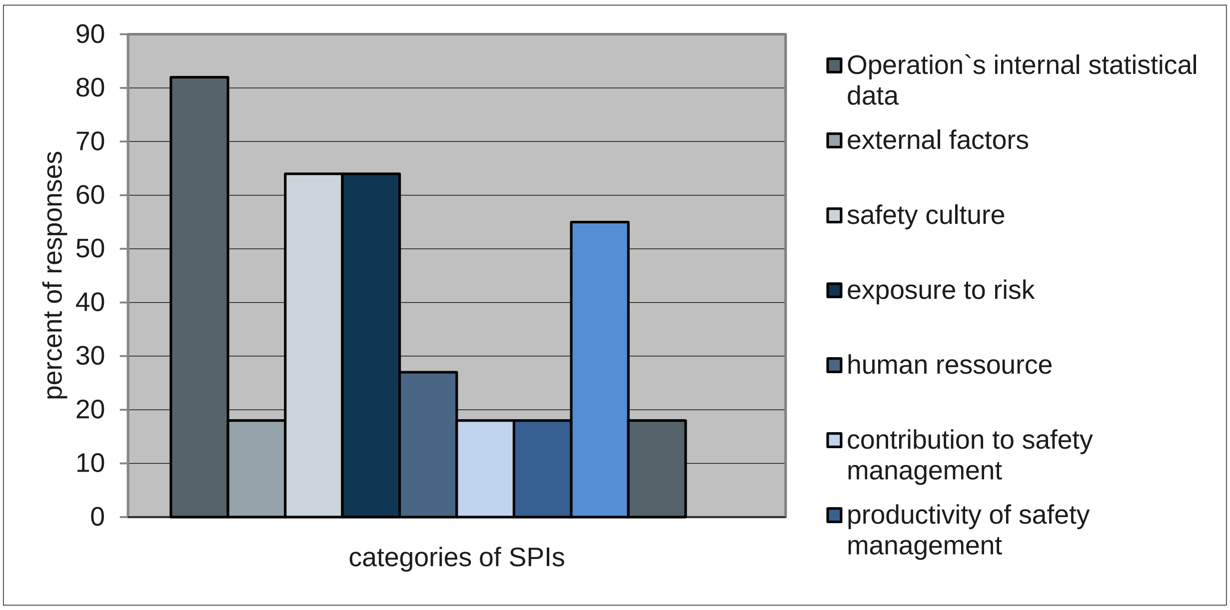


Figure 19: Categories of SPIs applied (own representation)

Another important evaluation to answer the third hypothesis about the necessity and comparability of safety performance indicators is the comparison of the currently applied SPIs with the cross company indicators that are considered useful for the organizations' safety management. The intention of these questions was to find out whether benchmarking is considered as useful for the organization or not and if the interest in the topics is conform. The figures in Table 5 listed under the column "Question A" are the responses to the question asking for the categories of indicators which are already applied at the respective organization and listed under column "Question B" are the figures which describe the question asking which cross company indicators would be beneficial for the respective safety management.

Evidently the categories concerning operations statistics data and exposure to risk data are those which are already applied and where the highest need of exchange exists. Especially the data gained through operations statistics are the easiest to compare. These are mostly pure numerical values which, however, have to be adjusted to the appropriate parameters such as the number of applications. The analysis showed the obvious that the same organizations having a demand on operations statistics data are those which use them already. This also applies to the exchange of human resources data. Although the percentage is lower (27%), the companies that use human resources data numbers and who have the needs for it are largely congruent. The still high percentage of demand for exposure risk data (64%) most likely lies in the need for basic numbers to establish exposure risk profiles, even though these numbers can't be easily transferred from one operation to the other.

There is no category that was particularly interesting for large or for small organizations. It was also not possible to find conformities between the categories and those companies that stated that all their SPIs gave numerical values and vice versa.

Table 5: Comparison of applied SPIs and indicators needed

answer options	Question A n	Question A %	Question B n	Question B %
Operation`s statistical data (e.g. mean time to failure, number of reports per year, serious incident rate, etc.)	9	**82%**	9	**82%**
External factors	2	**18%**	n/a	**n/a**
Safety culture (e.g. commitment, justness, information, Awareness, etc.)	7	**64%**	5	**45%**
Exposure to risk data (accepted missions in a certain risk profile, etc.)	7	**64%**	7	**64%**
Human resource data (staff turnover, postholder changes, etc.)	3	**27%**	3	**27%**
Contribution to safety management (over time)	2	**18%**	4	**36%**
Productivity of safety management (over time)	2	**18%**	4	**36%**
Effectiveness of safety management activity (over time)	6	**55%**	4	**36%**
Resilience indicators (e.g. number of staff with special competences)	2	**18%**	3	**27%**
others, for example:	0	**0%**	n/a	**n/a**

Safety culture and the effectiveness of safety management are the two categories where the demand for external data is less than the number of organizations that already use this data. Especially safety culture in particular is an area that can be very individual in an organization and depends on a large number of internal company parameters. It can be concluded from the low demand for exchange that the respondents are aware of the difficult applicability of the data concerning the safety culture and synonymous as above, it can also indicate an actually low use in the organization as stated. A bias for low interest in exchange could also be that this means that own sensitive operations data must be disclosed. Considering that safety is basically represented by a definition of a resilient system, it is surprising that resilience indicators represent the lowest percentages in the survey whereas the mandate for the exchange of safety management contribution and productivity is as high as expected.

While there are topics which are considered as not measureable by SPIs by the experts, the overall confidence in the application is high. Descriptions of the measurement can also be found in the literature for those areas that were not considered to be measurable like human factors or rare events. This can stand as a symbol for the development that still has to take place in the HEMS area.

6. Discussion

High-risk industries have always used methods based on the measurement of key indicators and using means such as historical data and prognostic models. Although these approaches are nothing new, major accidents in the past have shown all industries from nuclear power to the process and chemical industry and of course aviation, how important it is to give a structured background to the use of these models. Aviation, and especially the helicopter industry, can benefit from the increased use of models that are currently mostly only found at an academic level.

This work is intended to deal particularly with the use in the helicopter emergency medical services, since, as the survey also showed, this is an area which is currently on a threshold where state of the art safety management systems and the legislation require an application. A system level model was outlined, based on which the basic processes in the evaluation and monitoring of safety performance indicators are presented. Since there is no universal approach to the use of SPIs, operations can by means of this model plan a specific structure that meets the requirements of their particular organization.

While a common structure of SPIs in the field of HEMS can provide benefits on the exchange of data and therefore supports the efforts in the creation of risk based approaches, the consideration of system parameters that are measured by SPIs is always a philosophical question as well. What is the actual reality that is measured? SPI need to be reliable and valid in what they measure but even a reliable number will result in different interpretations. Therefore, the basic system structure and the meaning of the individual elements were dealt with in this paper and not the interpretation, which must then be done by the experts in an operation in an iterative process. It is a cyclical process in which the results flow repeatedly back into the various subsystems.

Although some experts in the survey feared excessive administrative effort with too little benefit in the usage of SPIs and also Wreathall (see Wreathall, 2009) questioned the use of categories in his paper, some applications could be shown here, that point out the advantages and necessity of using SPIs and their respective categories.

The weakness of the elements of the system level model as described is, that they only make sense in structural terms if applied completely. Individual elements can of course be implemented separately, but the overall outcome will most likely represent coincidences. In this sense, the elements of the system level model are minimum requirements for the structure. Janicak (2015) describes the shortcomings of safety performance indicators in a similar way, especially if measurements are not accurately assigned. Further he describes examples of shortcomings for missing definitions of measurements amongst other things:

- *A cause and effect relationship between the measurement and the activity does not exist.*
- *There are delays between the activity and the measurement.*
- *Reliability in the measurement process may be lacking.*

- *Outside factors may be influencing results, thus affecting the validity of the measurement.* (Janicak, 2015: 27)

The efficient use of SPIs is complex and therefore can presumably not be used by every company without restrictions. ICAO describes the management dilemma in the safety management manual as safety space, a metaphor for the sphere of action between the desired productivity and profitability on one side, versus the necessary protection through safety risk controls on the other side. The subject heading “production versus protection” describes the necessary balancing of resources between profitability and safety.

This paper was written in times of the corona crisis, which called for adaptations in the execution, especially in the empirical part. In addition to the Delphi analysis, expert interviews originally have been planned, which not only did not take place due to the restriction of personal contact options, but also because the safety staff of preclinical care facilities like HEMS operators had to deal with protective measures that changed almost every hour during this time. There was immense pressure on the security experts who, despite their tasks in this time of crisis, still found the time to pick up the questionnaires and thought processes of this study. Given this background, not only the return rate of the survey can be considered extraordinary, but also the interest in SPIs can be recognized.

Like any model, the approach with SPIs can only represent a part of reality. It will therefore need an intelligent link to other systems and further research. The novelty of the topic in this area of aviation also raises a number of questions that could not be investigated in this study but could serve as interesting research fields for further papers. For example, the effects of SPIs when they are initially introduced to a company, or the effects on individual categories compared with the respective outcome, for example according to the risk management cycle. As a result, universally valid guidelines for implementations could be developed. What would also be interesting is a research about the complexity of linking topics that a HEMS company deals with as a whole, for instance, an overlay of SPIs from patient care in the preclinical situation with the aviation and technical aspects of flight operations. Medicine is a similar safety sensible regime with completely different approaches – finding the interdependencies could be researched. A validation of the outcomes due to the use of SPIs in the field of aviation is still pending.

7. Conclusion

Across high-risk industries, there are several studies on the use of safety performance indicators but in the aviation sector only a few remain, and only one of them remains in the helicopter sector and not a single one treating the aspects of SPIs in the field of helicopter emergency medical services. Although there is a legal obligation to use SPIs, there is still a lot of work to do on the regulatory side like for example missing definitions for risk controls and acceptable risk levels need to be found for the industry. Not only through the literature research but also in the empirical approach, it was shown that there is on the user side still also no structured application in the helicopter emergency medical services. The current approach is primarily characterized by an isolated use of typically lagging indicators and those mostly being accident figures, which however allow little conclusion to be drawn about the current safety situation in an operation.
According to the system level model presented, the linking to a safety objective and a balanced relationship between leading and lagging indicators is desirable in order to get a realistic picture of the operation and to use prognostic methods based on the leading indicators as well as to use the lagging indicators to validate the use of the preventive measurements.
The multiscale natures with the interdependencies of SPIs make an application appear complex. Therefore, a categorization is expedient and it could be shown in this study that there are different uses of the categories. The research question regarding the relevance of the categories can be answered in that respect as the high-level categories are primarily functional for a cross-company exchange and scientific researches, whereas the more detailed categories are used in the direct application of the subsystems. The use of high level indicators for the cross-company exchange also ensures the comparability on an industry-wide basis.
The means to monitor SPIs vary from organization to organization as they strongly depend on the model and management system. There is nothing like a defined set of indicators appropriate to all organizations but a method using a risk map was presented to find categories, which characterize the interconnectedness of different organizational units. The found elements and links of this method can be grouped to clusters which will be measureable by the same means.
Safety Performance Indicators present a measured value at a given time and for a given situation, the overall context of this situation has to be established by the safety analyst. The representation of this overall context lies in the application of the system level model and in a comprehensible design of the parameters that are used. The requirements for monitoring and reporting, as well as the requirements for the allocation of resources will show which parameters in the model used are best suited to show a realistic picture of the safety in the operation in real time and with which risk influence factors have to be monitored.
An example of a representation has been shown in the chapter dealing with applications of SPIs where standard deviations are used and so creating a self-adapting system to exemplify the dynamic safety situation.
The research question about the challenges in defining safety performance indicators could be answered through the presentation of different methods to define SPIs.

Attention was also paid to the fact that not only complex scientific applications were presented, but considerations were made on how a basic set of SPIs can be found in the organization with simple means using the already existing methods like a bow tie analysis.

There is only a small number of organizations in Europe who provide HEMS and accordingly there are only a few experts who could be interviewed. For this reason, some of the statistics in this work deliberately pointed out the absolute numbers and therefore the small sample size of the Delphi analysis in order to refer to the weaknesses of this survey. Still some valid interpretations and trends can be drawn from, for instance a surprisingly high rate of black swan events throughout the industry or the fact that there is no connection between management systems and the application of SPIs.

The remaining question considering that there is nothing like a risk-free environment is how predictions can be made and how preventive measures can be used in a targeted manner. In the construction of a system like the one laid out in this paper, the proactive measures begin with the positioning of the SPIs and in the case of the corresponding outcome, this system level model, like all other approaches, becomes reactive because the measurement methods and the associated safety performance indicators have to be reasonably adapted.

List of references

Adams, J. (2007). Risk management: It's not rocket science... It's much more. *Risk Management, 54*, 36-40.

Adebiyi, K. A., Charles-Owaba, O. E., & Waheed, M. A. (2007). Safety performance evaluation models: a review. *Disaster Prevention and Management: An International Journal, 16*, 187. doi:10.1108/09653560710739504

AIZ. (1971). Abkommen über die internationale Zivilluftfahrt. Chicago, USA

Ale, B. J. M. (2005). Tolerable or Acceptable: A Comparison of Risk Regulation in the United Kingdom and in the Netherlands. *Risk Analysis, 25*, 241. doi:10.1111/j.1539-6924.2005.00585.x

Alessandro, V. (2011). Modelling dynamical processes in complex socio-technical systems. *Nature Physics, 8*. doi:10.1038/nphys2160

Amalberti, R. (2006). Optimum System Safety and
Optimum System Resilience: Agonistic or Antagonistic Concepts? In E. Hollnagel, D. D. Woods, & N. Leveson (Eds.), *Resilience engineering : : concepts and precepts* (xii, 397 pages) Aldershot, EnglandnBurlington, VT :: Ashgate.

Ammon, U. (2009). Delphi-Befragung. In S. Kühl, P. Strodtholz, & A. Taffertshofer (Eds.), *Handbuch Methoden der Organisationsforschung* (pp. 451). Wiesbaden: VS Verlag für Sozialwissenschaften.

ARMS Working Group. (2010). The ARMS methodology for operational risk assessment in aviation organisations. *London: Aviation Risk Management Solutions*.

Austrian Standards. (2014). ONR 49000–Risikomanagement für Organisationen und Systeme. *ON Osterreichisches Normungsinstitut Wien, 1*.

Badke-Schaub, P., Hofinger, G., & Lauche, K. (2012). *Human Factors: Psychologie sicheren Handelns in Risikobranchen* (2., überarbeitete Auflage ed.). Berlin, Heidelberg.

Bai, Y., & Jin, W.-L. (2016). Chapter 38 - Risk Assessment Methodology. In Y. Bai & W.-L. Jin (Eds.), *Marine Structural Design (Second Edition)* (pp. 709-723). Oxford: Butterworth-Heinemann.

Baram, M. (1998). Process safety management and the implications of organizational change. *Safety Management. The Challenge of Change. Oxford: Pergamon*.

BMK. (2020). Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie, Untersuchungsberichte Hubschrauber. Retrieved from https://www.bmk.gv.at/themen/sub/luftfahrt/berichte/hubschrauber.html access_date: 15.08.2020

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

Borg, I. (1996). *Facettentheorie*. Psychologische Rundschau, Münster

Bourrier, M. (1998). Elements for designing a self-correcting organisation: examples from nuclear power plants. *Safety Management, the challenge of change*. Pergamon

Canter, D. (1985). *Facet theory : approaches to social research*. New York, NY: Springer.

De Carvalho, P. V. R. (2011). The use of Functional Resonance Analysis Method (FRAM) in a mid-air collision to understand some characteristics of the air traffic management system resilience. *Reliability Engineering & System Safety, 96*(11), 1482-1498.

Di Gravio, G., Mancini, M., Patriarca, R., & Costantino, F. (2015). Overall safety performance of Air Traffic Management system: Forecasting and monitoring. *Safety Science, 72*, 351-362.

Drucker, P. (2009). If you can't measure it, you can't manage it. *MarketCulture Blog, 2015*. Retrieved from: https://blog.marketculture.com/2009/03/20/if-you-cant-measure-it-you-cant-manage-it-peter-drucker/ Access date: 15.08.2020

Duijm, N. J. (2015). Recommendations on the use and design of risk matrices. *Safety Science, 76*, 31. doi:10.1016/j.ssci.2015.02.014

EASA. (2016). *Practices for risk-based oversight*. Cologne

EASA. (2020a). EASA. Retrieved from https://www.easa.europa.eu/the-agency/the-agency Access date: 15.08.2020

EASA. (2020b). EASA Safety Review 2019. In. Cologne: European Union Aviation Safety Agency.

EC. (2019). 2018 road safety statistics: whats behind the figures? [Press release]. Retrieved from: https://ec.europa.eu/commission/presscorner/detail/en/MEMO_19_1990 Access date: 15.08.2020

EC. (Regulation 965/2012). Regulation (EU) No965/2012 of the European Parliament and of the Council. *Official Journal of the European Union*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32012R0965&qid=1589555891979&from=DE Access date: 15.08.2020

ERA. (2019). Safety Overview 2019 [Press release]. Retrieved from https://erail.era.europa.eu/documents/SPR.pdf Access date: 15.08.2020

European Aviation Safety Agency. (2007). *Annual report 2006*. Cologne.

Fahlbruch, B., & Schöbel, M. (2011). SOL – Safety through organizational learning: A method for event analysis. *Safety Science, 49*, 31. doi:10.1016/j.ssci.2010.05.004

Faulbaum, F., Prüfer, P., & Rexroth, M. (2009). *Was ist eine gute Frage?: Die systematische Evaluation der Fragenqualität* (Vol. 10). Wiesbaden.

Flight Safety Foundation. (2000). FSF ALAR briefing note 7.1—Stabilized approach. *Approach and landing accident reduction toolkit*.

Gerbec, M. (2013). Supporting organizational learning by comparing activities and outcomes of the safety-management system. *Journal of Loss Prevention in the Process Industries, 26*, 1127. doi:10.1016/j.jlp.2013.04.009

Gray, P. C. R., & Wiedemann, P. M. (1999). Risk management and sustainable development: mutual lessons from approaches to the use of indicators. *Journal of Risk Research, 2*, 218. doi:10.1080/136698799376808

Haberfellner, R., de Weck, O., Fricke, E., & Vössner, S. (2012). Systems Engineering–Grundlagen und Anwendung. 12. *Auflage. Zürich: Orell Füssli*, 978-3280040683

Häder, M. (2014). *Delphi-Befragungen : Ein Arbeitsbuch* (3. Aufl. 2014.. ed.) Wiesbaden :: Springer Fachmedien Wiesbaden : Imprint: Springer VS.

Hale, A. (2009). Why safety performance indicators? *Safety Science, 47*, 480. doi:10.1016/j.ssci.2008.07.018

Hale, A., Baram, M., & Hovden, J. (1998). Perspectives on safety management and change. *Safety Management–the Challenge of Change*, 1-18.

Hatry, H. P. (2006). *Performance measurement: Getting results* (2nd ed.): The Urban Insitute.

Hellevik, O. (1999). Research methodology in sociology and political science. *Norwegian: Forskiningsmetode i sosiologi og statsvitenskap). Universitetsforlaget, Oslo, Norway*.

Herrera, I. A., Hollnagel, E., & Håbrekke, S. (2010). *Proposing safety performance indicators for helicopter offshore on the Norwegian Continental Shelf*. Paper presented at the PSAM 10 - Tenth Conference on Probabilistic Safety Assessment and Management, Seattle, Wa, United States. https://hal-mines-paristech.archives-ouvertes.fr/hal-00613956 Access date: 25.08.2020

Hokstad, P., Jersin, E., & Sten, T. (2001). A risk influence model applied to North Sea helicopter transport. *Reliab. Eng. Syst. Saf., 74*, 322.

Hollnagel, E. (2012). *FRAM, the functional resonance analysis method: modelling complex socio-technical systems*: Ashgate Publishing, Ltd.

Hollnagel, E. (2014). *Safety-I and safety-II : : the past and future of safety management*. Farnham, Surrey, England Burlington, Vermont :: Ashgate

Holloway, J. (1995). *Performance measurement and evaluation* (1. publ.. ed.). London [u.a.]: Sage Publ.

Hopkins, A. (2009). Thinking About Process Safety Indicators. *Safety Science, 47*, 465. doi:10.1016/j.ssci.2007.12.006

Hug, T., Poscheschnik, G., Lederer, B., & Perzy, A. (2014). *Empirisch forschen* (überarb. Aufl. ed.).

IATA. (2019). *Safety Report 2018*. Montreal, Quebec: International Air Transport Association

ICAO. (1944). Convention on International Civil Aviation - Doc 7300.

ICAO. (2009). *Doc 9859: Safety Management Manual (SMM)*. (9292312952). Montreal, Canada: ICAO

ICAO. (2016). Annex 19 to the convention on international civil aviation, Safety Management. *2*.

ISO. (2014). Safety aspects - guidelines for the inclusion in standards, ISO/ICE guide 51:2104. In.

Janicak, C. A. (2015). *Safety metrics: Tools and techniques for measuring safety performance* (Vol. third edition). Maryland USA: Rowman & Littlefield Publishing Group, Inc.

Karanikas, N. (2015). 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*.

Kasper, S., Piric, S., de Boer, R., Roelen, A., & Karanikas, N. (2019). How does aviation industry measure safety performance? Current practice and limitations. *4*, 224-245. doi:10.1504/IJAM.2019.10019874

Khan, F. I. (2001). Use maximum-credible accident scenarios for realistic and reliable risk assessment. *Chemical engineering progress, 97*(11), 56-65.

Kjellén, U. (2009). The safety measurement problem revisited. *Safety Science, 47*, 489. doi:10.1016/j.ssci.2008.07.023

Kühl, S., Strodtholz, P., & Taffertshofer, A. (2009). *Handbuch Methoden der Organisationsforschung*: Springer.

Kurpiers, M. (2018). *Flight Safety Forum Minutes*. Retrieved from *Summary 2018*

Lališ, A. (2017). Time-series analysis and modelling to predict aviation safety performance index. *Transport Problems, 12*, 58. doi:10.20858/tp.2017.12.3.5

Lofquist, E. A. (2010). The art of measuring nothing: The paradox of measuring safety in a changing civil aviation industry using traditional safety metrics. *Safety Science, 48*, 1529. doi:10.1016/j.ssci.2010.05.006

Mickel, J. (2011). *Bewertung operationeller Flugrisiken*. Frankfurt am Main: Shaker Verlag

Montijn, C., & De Jong, H. (2006). Safety culture in air transport: Definition, characteristics, indicators and classification scheme. *NLR Memorandum ATSF-2006-150, NLR*.

Network of Analysts, S. P. I. S. G. (2014). EASA Member States Common Safety Performance Indicators. In EASA (Ed.). Colone.

ÖAMTC Flugrettung. (2020). *Jahresbericht 2019*. Wien

OECD. (2014). *Guidance on Developing Safety Performance Indicators For Industry*.

Øien, K. (2001). Risk Control of Offshore Installations–A Framework for the Establishment of Risk Indicators. *NTNU Report, 200104*.

Øien, K., Utne, I., & Herrera, I. (2011). Building Safety indicators: Part 1 – Theoretical foundation. *Safety Science, 49*, 161. doi:10.1016/j.ssci.2010.05.012

Perrow, C. B. (1994). *Accidents in high-risk systems*: National Emergency Training Center.

Pierce, F. D. (1995). Setting effective goals and objectives in safety and health programs. *Occupational Hazards, 57*(10), 169-173.

Reason, J. (2008). *Managing the risks of organizational accidents* (Reprint. ed.). Aldershot [u.a.]: Ashgate.

Roberts, K. H. (1989). New challenges in organizational research: high reliability organizations. *Industrial crisis quarterly, 3*(2), 111-125.

Rockwell, T. H. (1959). Safety performance measurement. *Journal of Industrial Engineering, 10*(1), 12-16.

Roelen, A., & Klompstra, M. (2012). The challenges in defining aviation safety performance indicators. *Helsinki, Finland. PSAM, 11*.

Roelen, A., Lin, P.-H., & Hale, A. (2011). Accident models and organisational factors in air transport: The need for multi-method models. *Safety Science, 49*, 10. doi:10.1016/j.ssci.2010.01.022

Roelen, A., van Aalst, R., Karanikas, N., Kaspers, S., Piric, S., & de Boer, R. J. (2018). Effectiveness of risk controls as indicator of safety performance. *AUP Advances, 1*, 189. doi:10.5117/ADV2018.1.ROEL

Seymore, P. (2018). *Helicopter Market Update*. Retrieved from https://www.iba.aero/insight/ibas-helicoper-webinar-slides-august-2019/ Access date: 30.04.2019

Shappell, S. A., & Wiegmann, D. A. (2001). Applying Reason: The human factors analysis and classification system (HFACS). *Human Factors and Aerospace Safety*.

Transport Canada. (2016). *Civil Aviation Integrated Management System Standard*. Ottawa, Ontario

Weick, K. E. (1987). Organizational Culture as a Source of High Reliability. *California Management Review, 29*, 127. doi:10.2307/41165243

Weick, K. E. (1990). The Vulnerable System: An Analysis of the Tenerife Air Disaster. *Journal of Management, 16*, 593. doi:10.1177/014920639001600304

Woudenberg, F. (1991). An evaluation of Delphi. *Technological Forecasting and Social Change, 40*(2), 131-150. doi:https://doi.org/10.1016/0040-1625(91)90002-W

Wreathall, J. (2009). Leading? Lagging? Whatever! *Saf. Sci., 47*, 494. doi:10.1016/j.ssci.2008.07.031

Appendix

Pretest data

For the pretest five persons from diverse professions were chosen to assess the test set-up:

<u>Test person 1:</u> Cpt. Dr. Pavel Müller is a doctor of law and as former president of EHAC very interested in HEMS safety

<u>Test person 2:</u> DI Bernhard Obholzer is, next to his job as architect, a technical crewmember on a rescue helicopter and part of the safety team of the ÖAMTC air ambulance

<u>Test person 3:</u> Cpt. Markus Seidl, M.eng. is a rescue helicopter pilot and deputy safety manager in the ÖAMTC air ambulance

<u>Test person 4:</u> Mag.(FH) Bernd Lang is sociologist and compliance monitoring manager in the ÖAMTC air ambulance

<u>Test person 5:</u> Dr. Christian Dopler is an anesthesiologist, serves as medical doctor on a rescue helicopter and is part of the safety team of the ÖAMTC air ambulance

<table>
<tr><th>PRETEST</th><th>P 1</th><th>P 2</th><th>P 3</th><th>P 4</th><th>P 5</th></tr>
<tr><td>Were the questions understandable?</td><td>yes</td><td>yes</td><td>yes</td><td rowspan="6">answers were too late to be considered for the testing</td><td>yes</td></tr>
<tr><td>Where there any problems under-standing the task?</td><td>no</td><td>no, but needed additional information</td><td>no</td><td>no</td></tr>
<tr><td>Was the survey too long or too short (e.g. could the interest be kept up throughout the interrogation)?</td><td>ok</td><td>ok</td><td>ok</td><td>ok</td></tr>
<tr><td>Did any questions raise discomfort (e.g through talking about confidential information)?</td><td>no</td><td>no; third option if it´s not practi-cable to answer</td><td>no</td><td>no</td></tr>
<tr><td>Did any technical problems arise?</td><td>no</td><td>2 attempts; due to bad mobile connection</td><td>no</td><td>no</td></tr>
<tr><td>Any additional comments?</td><td>no</td><td>third option in answers</td><td>third option in answers</td><td>no</td></tr>
</table>

Codebook of the first round of the survey

In order to identify the attributes and to relate the questions to the content, the possible answers have been coded with variables which are listed in the following codebook. The deduced correlations found can be reproduced using the code book.

Project-ID	*50461*
URL of the survey	*https://ww2.efs-survey.at/uc/SPI/*
date	*15.06.2020 07:17:58*

Valuation (Generally and basic information for all three hypothesis) (PGID 188519)

How many helicopters are currently active in your operation? (q_258055 - Typ 141)

v_1	v_1	varchar	How many helicopters are in your operation?

How many flight missions does the operation you work for achieve on average during one year? (q_258057 - Typ 111)

v_8	v_8	int	How many flight missions does the operation you work for achieve on average during one year?
		1	*< 1.000*
		2	*1.000 - 5.000*
		3	*5.001 - 10.000*
		4	*10.001 - 15.000*
		5	*15.001 - 20.000*
		6	*20.001 - 25.000*
		7	*> 25.000*

How many employees are currently in the operation you work for? (q_258059 - Typ 141)

v_9	v_9	varchar	How many employees are currently in the operation you work for?

Do you monitor safety in the operation you work for? (q_258060 - Typ 121)

v_10	v_10	int	through SPIs
		0	*not quoted*
		1	*quoted*
v_11	v_11	int	regular reviews
		0	*not quoted*
		1	*quoted*
v_12	v_12	int	audits

		0	*not quoted*
		1	*quoted*
v_13	v_13	int	surveys
		0	*not quoted*
		1	*quoted*
v_15	v_15	int	Others, for example:
		0	*not quoted*
		1	*quoted*
v_16	v_16	varchar	Others, for example:

Do you use an integrated management system? (q_258061 - Typ 111)

v_17	v_17	int	Do you use an integrated management system?
		1	*Yes*
		2	*No*

Are compliance and safety two separated divisions? (q_258062 - Typ 111)

v_18	v_18	int	Are compliance and safety two separated divisions?
		1	*Yes*
		2	*No*

Does senior management in your operation demand SPIs as a reporting method? (q_258063 - Typ 111)

v_19	v_19	int	Does senior management in your operation demand SPIs as a reporting method?
		1	*Yes*
		2	*No*

What is your evaluation of the "severity" in your risk assessments based on? (q_258064 - Typ 121)

v_20	v_20	int	past events in the company
		0	*not quoted*
		1	*quoted*
v_21	v_21	int	events in the industry
		0	*not quoted*
		1	*quoted*
v_22	v_22	int	exposition to risk
		0	*not quoted*
		1	*quoted*
v_23	v_23	int	probabilistic models
		0	*not quoted*
		1	*quoted*
v_25	v_25	int	others, for example:
		0	*not quoted*

		1	*quoted*
v_26	v_26	varchar	others, for example:

How do you consider the importance of the application of SPIs for operational safety? (q_258065 - Typ 111)

v_27	v_27	int	How do you consider the importance of the application of SPIs for operational safety?
		1	*1 - No importance at all*
		2	*2*
		3	*3*
		4	*4*
		5	*5 - Very high importance*

By what means can operational safety be tracked best? (q_258066 - Typ 142)

v_28	v_28	blob	By what means can operational safety be tracked best?

Valuation (General and basic information for all three hypothesis) – Follow up question (PGID 188520)

On average how often are the monitoring systems for safety used in your operation? (q_258067 - Typ 311)

v_85	v_85	int	List item: through SPIs (Source variable: v_10 (through SPIs))
		1	*less than once per month*
		2	*two to five times per month*
		3	*six to ten times per month*
		4	*more than ten times per month*
		5	*never*
v_86	v_86	int	List item: regular reviews Source variable: v_11 (regular reviews))
		1	*less than once per month*
		2	*two to five times per month*
		3	*six to ten times per month*
		4	*more than ten times per month*
		5	*never*
v_87	v_87	int	List item: audits (Source variable: v_12 (audits))
		1	*less than once per month*
		2	*two to five times per month*
		3	*six to ten times per month*
		4	*more than ten times per month*
		5	*never*
v_88	v_88	int	List item: surveys (Source variable: v_13 (surveys))
		1	*less than once per month*

		2	*two to five times per month*
		3	*six to ten times per month*
		4	*more than ten times per month*
		5	*never*
v_89	v_89	int	List item: #v_16# (Source variable: v_15 (Others, for example:))
		1	*less than once per month*
		2	*two to five times per month*
		3	*six to ten times per month*
		4	*more than ten times per month*
		5	*never*

Categorization (Hypothesis 1 a, b) (PGID 188521)

Do you assess safety culture in your organization using SPIs? If yes how? (q_258068 - Typ 111)

v_39	v_39	int	Do you assess safety culture in your organization using SPIs? If yes how?
		1	*Yes, for example:*
		2	*No*
v_40	v_40	varchar	Yes, for example:

Do you assess change management in your organization using SPIs? (q_258069 - Typ 111)

v_41	v_41	int	Do you assess change management in your organization using SPIs?
		1	*Yes*
		2	*No*
		3	*We assess change management, but we don't use SPIs*

Does your safety management system provide categories of SPIs? (q_258070 - Typ 111)

v_42	v_42	int	Does your safety management system provide categories of SPIs?
		1	*Yes*
		2	*No*

Which of the following categories of SPIs does the organization you work for apply? (q_258072 - Typ 121)

v_44	v_44	int	Operation’s internal statistical data
		0	*not quoted*
		1	*quoted*
v_45	v_45	int	external factors
		0	*not quoted*

		1	*quoted*
v_46	v_46	int	safety culture
		0	*not quoted*
		1	*quoted*
v_47	v_47	int	exposure to risk
		0	*not quoted*
		1	*quoted*
v_48	v_48	int	human resource
		0	*not quoted*
		1	*quoted*
v_49	v_49	int	contribution to safety management
		0	*not quoted*
		1	*quoted*
v_50	v_50	int	productivity of safety management
		0	*not quoted*
		1	*quoted*
v_51	v_51	int	effectiveness of safety activity
		0	*not quoted*
		1	*quoted*
v_52	v_52	int	resilience indicators
		0	*not quoted*
		1	*quoted*
v_53	v_53	int	others, for example:
		0	*not quoted*
		1	*quoted*
v_54	v_54	varchar	others, for example:

Do your categories of SPIs (grouped SPIs) provide numerical values? (q_258073 - Typ 111)

v_55	v_55	int	Do your categories of SPIs (grouped SPIs) provide numerical values?
		1	*Yes*
		2	*No*

Are there categories in use which you consider as not relevant for operational safety? (e.g. financial aspects) (q_258074 - Typ 111)

v_56	v_56	int	Are there categories in use which you consider as not relevant for operational safety? (e.g. financial aspects)
		6	*Yes, for example:*
		7	*No*
v_57	v_57	varchar	Yes, for example:

Impact on operations (Hypothesis 2 a) (PGID 188522)

Is there a distribution of safety information to operational employees using SPIs? (q_258075 - Typ 111)

v_58	v_58	int	Is there a distribution of safety information to operational employees using SPIs?
		1	*Yes*
		2	*No*
		3	*There is a distribution of safety information, but without using SPIs*

Does your safety management system utilize a monitoring system for preventive safety activities? (q_258076 - Typ 111)

v_59	v_59	int	Does your safety management system utilize a monitoring system for preventive safety activities?
		1	*Yes*
		2	*No*

Are there areas in your operation which cannot be measured by SPIs? If yes, please cite: (q_258077 - Typ 111)

v_60	v_60	int	Are there areas in your operation which cannot be measured by SPIs? If yes, please cite:
		1	*Yes, for example:*
		2	*No*
v_61	v_61	varchar	Yes, for example:

Does your organization apply SPIs which consider an accumulation of events? If yes please name the SPIs. (q_258078 - Typ 111)

v_62	v_62	int	Does your organization apply SPIs which consider an accumulation of events? If yes please name the SPIs:
		1	*Yes, for example:*
		2	*No*
v_63	v_63	varchar	Yes, for example:

Does your safety management system compare findings from incident investigations with the monitoring of preventive activities? (q_258079 - Typ 111)

v_64	v_64	int	Does your safety management system compare findings from incident investigations with the monitoring of preventive activities?
		1	*Yes*
		2	*No*

Have you heard of the following methods and are they used in your organization: Systems-theoretic accident modelling and processes (STAMP); Functional resonance analysis method (FRAM); Risk influence modeling (RIF); Human factors analysis and classification system (HFACS)? (q_258080 - Typ 311)

v_65	v_65	int	STAMP
		1	*I have heard from this method*
		2	*No, I haven't heard from that method*
		3	*This method is applied in the company I work for*
v_66	v_66	int	FRAM
		1	*I have heard from this method*
		2	*No, I haven't heard from that method*
		3	*This method is applied in the company I work for*
v_67	v_67	int	RIF
		1	*I have heard from this method*
		2	*No, I haven't heard from that method*
		3	*This method is applied in the company I work for*
v_68	v_68	int	HFACS
		1	*I have heard from this method*
		2	*No, I haven't heard from that method*
		3	*This method is applied in the company I work for*

challenge in def. / comparable (Hyp. 3 a, b) (PGID 188523)

Does you operation keep track of specific values like time between component failures? (q_258081 - Typ 111)

v_70	v_70	int	Does you operation keep track of specific values like time between component failures?
		1	*Yes*
		2	*No*

Does your operation keep track of specific values like mean time between failures? (q_258089 - Typ 111)

v_71	v_71	int	Does your operation keep track of specific values like mean time between failures?
		1	*Yes*
		2	*No*

Do all of your SPIs provide numerical values like numbers or ratios? (q_258090 - Typ 111)

v_72	v_72	int	Do all of your SPIs provide numerical values or number or ratios?
		1	*Yes*
		2	*No*

How often are your SPIs updated? (q_258125 - Typ 111)

v_90	v_90	int	How often are your SPIs updated?
		1	*ad hoc*
		2	*weekly*
		3	*monthly*
		4	*bi annual*
		5	*annually*

Does your organization use external indicators (e.g. parameters from authorities) for comparison? If yes, please provide which information is used. (q_258093 - Typ 111)

v_79	v_79	int	Does your organization use external indicators (e.g. parameters from authorities) for comparison? If yes, please provide which information is used:
		1	*Yes, for example:*
		2	*No*
v_84	v_84	varchar	Yes, for example:

Does your operation use SPIs concerning organizational culture and policies? (q_258094 - Typ 111)

v_80	v_80	int	Does your operation use SPIs concerning organizational culture and policies?
		1	*Yes*
		2	*No*

Are there any cross company indicators that would help your safety management? (q_258126 - Typ 121)

v_92	v_92	int	Operation's statistical data (e.g. mean time to failure, number of reports per year, serious incident rate, etc.)
		0	*not quoted*
		1	*quoted*
v_93	v_93	int	safety culture (e.g. commitment, justness, information, awareness, etc.)
		0	*not quoted*
		1	*quoted*
v_94	v_94	int	exposure to risk data (accepted missions in a certain risk profile, etc.)
		0	*not quoted*
		1	*quoted*
v_95	v_95	int	human resource data (staff turnover, postholder changes, etc.)
		0	*not quoted*
		1	*quoted*
v_96	v_96	int	contribution to safety management (over time)

		0	*not quoted*
		1	*quoted*
v_97	v_97	int	productivity of safety management (over time)
		0	*not quoted*
		1	*quoted*
v_98	v_98	int	effectiveness of safety management activity (over time)
		0	*not quoted*
		1	*quoted*
v_99	v_99	int	resilience indicators (e.g. number of staff with special competences)
		0	*not quoted*
		1	*quoted*

Admin Questions (PGID 188524)

In which country are the headquarters of your operation located? (q_258096 - Typ 141)

v_82	v_82	varchar	In which country are the headquarters of your operation located?

Comments (PGID 188525)

Please provide an e-mail address where the evaluation of this survey and the related questions for the abstract can be sent to: (q_258507 - Typ 141)

v_100	v_100	varchar (including check: mail-address)	E-Mail contact

Are there any additional comments you want us to consider for the follow-up examination? Is there anything you would like to address? (q_258097 - Typ 142)

v_83	v_83	blob	Are there any further comments, anything you would like to address?

Results of the first round of the survey

The results of the survey are shown below. The question is on the top line. If there were several possible answers, they are shown on the second line. The third line contains the variables of the respective answers as they are referenced in the code book. “Quoted” means yes – e.g. the questioned parameter is monitored, while “not quoted” means the opposite.
The query tool also determines technical data such as the duration of the completion or the browser used. This data was hidden for the sake of better legibility, as were the answers to contact details and the country of the head office, since publication of this data would compromise the anonymity of the survey. Out of the same reason due to data protection and protection of company internal information which was shared during the survey, the names of the participating organizations are not published.

The second round of the survey is structured analogously.

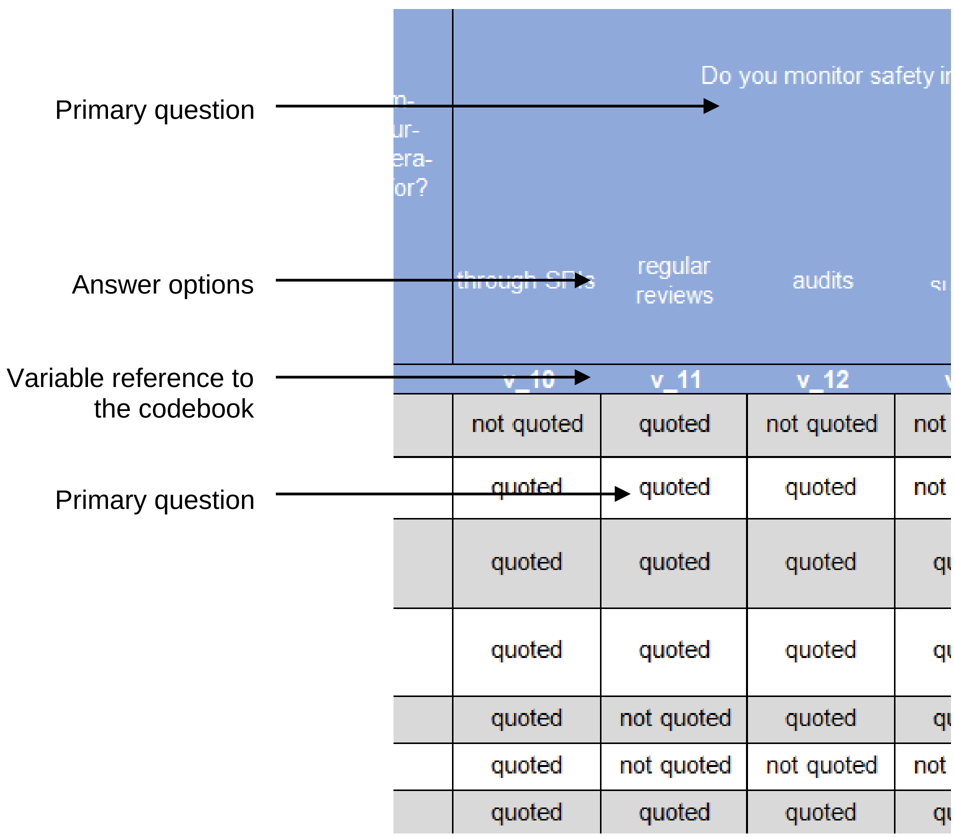


Figure 20: Structure of the results of the survey

How many helicopters are currently active in your opera-tion?	How many flight missions does the operation you work for achieve on average during one year?	How many employees are currently in the operation you work for?	Do you monitor safety in the operation you work for?				Do you monitor safety in the operation you work for?		Do you use an integrated management system?
			through SPIs	regular reviews	audits	surveys	Others, for example:	Others, for example:	
v_1	v_8	v_9	v_10	v_11	v_12	v_13	v_15	v_16	v_17
3	1.000 - 5.000	8	not quoted	quoted	not quoted	not quoted	not quoted	-99	Yes
3	1.000 - 5.000	75	quoted	quoted	quoted	not quoted	not quoted	-99	Yes
9	1.000 - 5.000	126	quoted	quoted	quoted	quoted	not quoted	-99	Yes
20	15.001 - 20.000	370	quoted	quoted	quoted	quoted	quoted	Reporting system, investigations, risk assessments	Yes
90	> 25.000	900	quoted	not quoted	quoted	quoted	not quoted	-99	No
ca. 50	> 25.000	ca. 1200	quoted	not quoted	not quoted	not quoted	not quoted	-99	No
17	10.001 - 15.000	250	quoted	quoted	quoted	quoted	not quoted	-99	Yes
9	1.000 - 5.000	150	quoted	quoted	quoted	not quoted	quoted	Inspections	Yes
2	< 1.000	40	quoted	quoted	quoted	quoted	quoted	Reports	Yes
2	< 1.000	30	not quoted	quoted	quoted	not quoted	not quoted	-99	Yes
54	> 25.000	250	not quoted	quoted	quoted	not quoted	not quoted	-99	Yes

Are compliance and safety two separate divi-sions?	Does senior management in your operation demand SPIs as a reporting method?	What is your evaluation of the "severity" in your risk assessments based on?		What is your evaluation of the "severity" in your risk assessments based on?				How do you consider the importance of the application of SPIs for opera-tional safety?
		past events in the company	events in the industry	exposition to risk	probabilistic models	others, for example:	others, for example:	
v_18	v_19	v_20	v_21	v_22	v_23	v_25	v_26	v_27
Yes	Yes	not quoted	quoted	not quoted	not quoted	not quoted	-99	3
Yes	Yes	quoted	not quoted	quoted	quoted	not quoted	-99	4
No	Yes	quoted	not quoted	quoted	quoted	not quoted	-99	5 - Very high importance
Yes	Yes	quoted	quoted	quoted	quoted	not quoted	-99	5 - Very high importance
Yes	No	quoted	quoted	not quoted	not quoted	not quoted	-99	4
Yes	Yes	quoted	not quoted	not quoted	not quoted	not quoted	-99	5 - Very high importance
Yes	Yes	quoted	quoted	quoted	quoted	not quoted	-99	3
No	Yes	not quoted	not quoted	not quoted	quoted	not quoted	-99	4
Yes	Yes	not quoted	not quoted	not quoted	quoted	not quoted	-99	4
Yes	No	not quoted	not quoted	quoted	quoted	not quoted	-99	4
Yes	No	quoted	not quoted	quoted	not quoted	not quoted	-99	4

By what means can operational safety be tracked best?
v_28
Through debriefing and Teamwork
Evaluation of Operation, Risk mitigation, Training of all personnel and risk warning; Incident analysis & broaden your horizons (peek over the neighbor's fence)
Best tracked? With enough safety staff. The company (and the authority) accepted the just culture reporting system. The crew accepted the managers. The crew accept and apply the reporting policy and system. The safety staff must make and record all safety events. The responsible person should periodically check it, summarize it, and evaluate the events. And the nominated person who made a corrective action should be based on the information.
Monitoring the efficiency of barriers established to prevent the most probable accident causes
money which has been spent different than planned for risk controls / mitigating actions
SPIs, Audits
By safety reports and by line checks.
Occurrence reports, Deviations from SOPs, Crew discretion reports, SACA/SAFA reports, Inspections, MEL...
SPIs, trends.
Reports
SPI / KPI

On average how often are the monitoring systems for safety used in your operation?					Do you assess safety culture in your organization using SPIs? If yes how?		Do you assess change management in your organization using SPIs?
through SPIs	regular reviews	audits	surveys	Others, for example:		Yes, for example:	
v_85	v_86	v_87	v_88	v_89	v_39	v_40	v_41
-77	less than once per month	-77	-77	-77	No	-99	Yes
two to five times per month	two to five times per month	two to five times per month	-77	-77	Yes, for example:	HHO, HEMS, NVG, CM & QM	Yes
less than once per month	two to five times per month	two to five times per month	less than once per month	-77	Yes, for example:	Mandatory and voluntary reporting number per 1000 fh, Mandatory and voluntary reporting number per 1000 landings	Yes
two to five times per month	less than once per month	two to five times per month	less than once per month	more than ten times per month	Yes, for example:	Monitoring number of reports submitted for contributions to risk management	Yes
less than once per month	-77	two to five times per month	less than once per month	-77	Yes, for example:	Safety Surveys, use of voluntary flight safety reports	We assess change management, but we don't use SPIs
less than once per month	-77	-77	-77	-77	No	-99	No
less than once per month	two to five times per month	two to five times per month	six to ten times per month	-77	No	-99	We assess change management, but we don't use SPIs
less than once per month	more than ten times per month	less than once per month	-77	six to ten times per month	Yes, for example:	NP's management meetings	Yes
less than once per month	two to five times per month	two to five times per month	less than once per month	more than ten times per month	No	-99	We assess change management, but we don't use SPIs
-77	less than once per month	less than once per month	-77	-77	No	-99	No
-77	less than once per month	less than once per month	-77	-77	No	-99	We assess change management, but we don't use SPIs

Does your safety management system provide categories of SPIs?	Which of the following categories of SPIs does the organization you work for apply?				Which of the following categories of SPIs does the organization you work for apply?				
	Operation`s internal tistical data	external factors	safety culture	exposure to risk sta-	man resource	contribution to safety management	productivity of safety management	effectiveness of safety activity	resilience hu- indicators
v_42	v_44	v_45	v_46	v_47	v_48	v_49	v_50	v_51	v_52
Yes	not quoted	not quoted	quoted	not quoted	not quoted	not quoted	not quoted	not quoted	not quoted
Yes	quoted	quoted	quoted	quoted	quoted	not quoted	quoted	quoted	not quoted
Yes	quoted	quoted	quoted	quoted	quoted	not quoted	not quoted	quoted	not quoted
Yes	quoted	not quoted	quoted	quoted	quoted	quoted	not quoted	quoted	quoted
Yes	quoted	not quoted	quoted	quoted	not quoted	not quoted	not quoted	not quoted	not quoted
Yes	quoted	not quoted	not quoted	quoted	not quoted	not quoted	quoted	not quoted	not quoted
No	quoted	not quoted	not quoted	not quoted	not quoted	not quoted	not quoted	not quoted	not quoted
Yes	quoted	not quoted	quoted	quoted	not quoted	quoted	not quoted	quoted	not quoted
Yes	quoted	not quoted	not quoted	not quoted	not quoted	not quoted	not quoted	quoted	quoted
No	not quoted	not quoted	not quoted	quoted	not quoted	not quoted	not quoted	not quoted	not quoted
No	quoted	not quoted	quoted	not quoted	not quoted	not quoted	not quoted	quoted	not quoted

Which of the following categories of SPIs does the organization you work for apply? others, for example:	Do your categories of SPIs (grouped SPIs) provide numerical values?	Are there categories in use which you consider as not relevant for operational safety? (e.g. financial aspects) Yes, for example:		Is there a distribution of safety information to operational employees using SPIs?	Does your safety management system utilize a monitoring system for preventive safety activities?	Are there areas in your operation which cannot be measured by SPIs? If yes, please cite:
v_54	v_55	v_56	v_57	v_58	v_59	v_60
-99	No	No	-99	Yes	No	No
-99	Yes	No	-99	There is a distribution of safety information, but without using SPIs	Yes	No
-99	Yes	No	-99	Yes	Yes	Yes, for example:
-99	Yes	No	-99	Yes	Yes	Yes, for example:
-99	No	No	-99	Yes	Yes	Yes, for example:
-99	No	Yes, for example:	Patient Safety, Financial, Occupational Safety and Health, Environmental security	There is a distribution of safety information, but without using SPIs	Yes	No
-99	Yes	No	-99	There is a distribution of safety information, but without using SPIs	No	Yes, for example:
-99	Yes	No	-99	Yes	Yes	No
-99	Yes	No	-99	Yes	Yes	No
-99	No	No	-99	There is a distribution of safety information, but without using SPIs	Yes	Yes, for example:
-99	Yes	No	-99	There is a distribution of safety information, but without using SPIs	No	Yes, for example:

Are there areas in your operation which cannot be measured by SPIs? If yes, please cite: Yes, for example:	Does your organization apply SPIs which consider an accumulation of events? If yes please name the SPIs.	Does your organization apply SPIs which consider an accumulation of events? If yes please name the SPIs. Yes, for example:	Does your safety management system compare findings from incident investigations with the monitoring of preventive activities?	Have you heard of the following methods and are they used in your organization: Systems-theoretic accident modelling and processes? STAMP
v_61	v_62	v_63	v_64	v_65
-99	No	-99	Yes	I have heard of this method
-99	Yes, for example:	Technical Certification, Design, attention and distraction, Airspace, Operation (HHO & HEMS), weather, Fueling, MX-Activity, HR himself ... and so on	No	I have heard of this method
The medical areas a separated area	Yes, for example:	severity of the event	Yes	I have heard of this method
A selection of the vital SPIs is considered best, keeping the number of SPIs to ALARP	Yes, for example:	Air traffic conflicts, navigational errors, HEC, Air space violations, Technical affecting flight safety, Fatigue, weather issues	Yes	No, I haven't heard of that method
actually no HFDM in use	No	-99	Yes	-99
-99	Yes, for example:	Airprox, Birdstrike, Icing, aircraft limit exceedance, incident due to rotor downwash	No	No, I haven't heard of that method
Safety feedbacks from crews, findings from line checks.	Yes, for example:	Damage to others	No	No, I haven't heard of that method
-99	Yes, for example:	Third Party Damage - SAFA/SACA - Crew discretion reports..	Yes	No, I haven't heard of that method
-99	Yes, for example:	Engine power check failures. Go arounds from site	Yes	No, I haven't heard of that method
-99	No	-99	Yes	No, I haven't heard of that method
financial	No	-99	Yes	No, I haven't heard of that method

Have you heard of the following methods and are they used in your organization: Systems-theoretic accident modelling and processes (STAMP); Functional resonance analysis method (FRAM); Risk influence modeling (RIF); Human factors analysis and classification system (HFACS)? FRAM	RIF	HFACS	Does you operation keep track of specific values like time between component failures?	Does you operation keep track of specific values like mean time between failures?	Do all of your SPIs provide numerical values like numbers or ratios?	How often are your SPIs updated?
v_66	v_67	v_68	v_70	v_71	v_72	v_90
No, I haven't heard of that method	I have heard of this method	I have heard of this method	Yes	No	No	annually
No, I haven't heard of that method	I have heard of this method	I have heard of this method	Yes	Yes	Yes	monthly
No, I haven't heard of that method	No, I haven't heard of that method	This method is applied in the company I work for	No	Yes	Yes	weekly
No, I haven't heard of that method	No, I haven't heard of that method	I have heard of this method	Yes	Yes	Yes	monthly
-99	-99	This method is applied in the company I work for	Yes	Yes	Yes	ad hoc
No, I haven't heard of that method	No, I haven't heard of that method	This method is applied in the company I work for	No	No	Yes	ad hoc
No, I haven't heard of that method	No, I haven't heard of that method	I have heard of this method	No	No	No	annually
No, I haven't heard of that method	No, I haven't heard of that method	No, I haven't heard of that method	Yes	Yes	Yes	ad hoc
I have heard of this method	No, I haven't heard of that method	I have heard of this method	Yes	Yes	Yes	monthly
No, I haven't heard of that method	No, I haven't heard of that method	No, I haven't heard of that method	Yes	Yes	No	annually
No, I haven't heard of that method	I have heard of this method	This method is applied in the company I work for	Yes	Yes	Yes	ad hoc

Does your organization use external indicators (e.g. parameters from authorities) for comparison? If yes, please provide which information is used.	Does your organization use external indicators (e.g. parameters from authorities) for comparison? If yes, please provide which information is used. Yes, for example:	Does your operation use SPIs concerning organizational culture and policies?	Are there any cross-company indicators, that would help your safety management? Operation`s statistical data (e.g. mean time to failure, number of reports per year, serious incident rate, etc.)	Are there any cross-company indicators that would help your safety management? safety culture (e.g. commitment, accepted justness, information, awareness, risk profile, etc.)	exposure to risk data (actual missions in a certain etc.)
v_79	v_84	v_80	v_92	v_93	v_ 94
No	-99	Yes	not quoted	quoted	not quoted
Yes, for example:	Mother company, TCDS	Yes	quoted	quoted	quoted
No	-99	Yes	quoted	quoted	quoted
No	-99	No	quoted	quoted	quoted
No	-99	Yes	quoted	not quoted	quoted
No	-99	Yes	quoted	not quoted	not quoted
Yes, for example:	Information of the target level of safety from other helicopter operators.	No	quoted	not quoted	not quoted
No	-99	No	quoted	not quoted	quoted
No	-99	Yes	not quoted	not quoted	not quoted
No	-99	No	quoted	quoted	quoted
No	-99	No	quoted	not quoted	quoted

Are there any cross company indicators that would help your safety management?				
human resource data (staff turnover, postholder changes, etc.)	contribution to safety management (over time)	productivity of safety management (over time)	effectiveness of safety management activity (over time)	resilience indicators (e.g. number of staff with special competences)
v_95	v_96	v_97	v_98	v_99
not quoted	not quoted	not quoted	not quoted	not quoted
quoted	quoted	not quoted	quoted	quoted
not quoted	not quoted	not quoted	not quoted	not quoted
quoted	quoted	quoted	quoted	quoted
not quoted	quoted	quoted	not quoted	not quoted
not quoted	not quoted	quoted	not quoted	not quoted
not quoted	not quoted	not quoted	not quoted	not quoted
quoted	not quoted	not quoted	not quoted	not quoted
not quoted	quoted	quoted	quoted	quoted
not quoted	not quoted	not quoted	not quoted	not quoted
not quoted	not quoted	not quoted	quoted	not quoted

Codebook of the second round of the survey

Projekt-ID	*50603*
URL of the survey	*https://ww2.efs-survey.at/uc/SPI_2/*
Datum	*08.07.2020 08:59:47*

Valuation (Generally and basic information for all three hypothesis) (PGID 189840)

The number of employees of the participating organizations range from less than 8 up to several hundred. This also implies different efforts in the safety management. How many full-time equivalent employees are currently engaged in your safety management? (q_260563 - Typ 111)

v_101	v_1_1	int	The number of employees of the participating organizations range from less than 8 up to several hundred. This also implies different efforts in the safety management. How many full-time equivalent employees are currently engaged in your safety management
		1	*Only one part-time employee*
		2	*1 full-time employee*
		3	*2 full-time employees*
		4	*3 full-time employees*
		5	*4 full-time employees*
		6	*5 full-time employees*
		7	*6 or more full-time employees*

82% of the respondents of the first round stated that they consider the importance of the application of SPIs as high or very high. But only 73% actually use SPIs in their operation. Does your organization consider changing the amount of usage of SPIs in the future? (q_260564 - Typ 111)

v_102	v_2_1	int	82% of the respondents of the first round stated that they consider the importance of the application of SPIs as high or very high. But only 73% actually use SPIs in their operation. Does your organization consider to change the amount of usage of SPIs in
		1	*Yes, more SPIs than now, for example:*
		2	*No change in the amount is planned*
		3	*yes, less SPIs than now, for example:*
v_103	v_2_2	varchar	Yes, more SPIs than now, for example:
v_104	v_2_3	varchar	yes, less SPIs than now, for example:

In your opinion, what are the main reasons for using or not using SPIs? (q_260565 - Typ 142)

v_105	v_3_1	blob	In your opinion, what are the main reasons for using or not using SPIs?

Categorization (Hypothesis 1 a, b) (PGID 189841)

Following the ICAOs approach on progress in aviation safety, a disaggregation of certain factors helps with the allocation in development stages of a safety system. Does your organization track any of the following influences using SPIs? (q_260566 - Typ 121)

v_113	v_4_1	int	technical (technical and technological failures)
		0	*not quoted*
		1	*quoted*
v_114	v_4_2	int	human factors (including man/machine interface errors)
		0	*not quoted*
		1	*quoted*
v_115	v_4_3	int	organizational (systemic and culture impacts)
		0	*not quoted*
		1	*quoted*
v_116	v_4_4	int	total system (failures due to the complexity of the system e.g. interfaces of organizations or influence of governmental administration)
		0	*not quoted*
		1	*quoted*
v_111	v_4_5	int	Others, for example:
		0	*not quoted*
		1	*quoted*
v_112	v_4_6	varchar	Others, for example:
v_152	v_4_7	int	none
		0	*not quoted*
		1	*quoted*

SPIs can either be reactive and therefore based on historical data (like incidents in the company), or proactive trying to disclose situations with higher or evolving risk. How many of the SPIs used in your safety management can be considered as proactive? (q_260567 - Typ 911)

v_117	v_5_1	int	Antwortoption

Impact on operations (Hypothesis 2 a) (PGID 189842)

Most organizations of the survey conduct between two to five audits per month, which could be interpreted that audits seem to be accepted as a preventive tool. Do the audits used in your organization provide numerical values, that can be interpreted as SPIs? (q_260571 - Typ 111)

v_123	v_6_1	int	Most organizations of the survey conduct between two to five audits per month, which could be interpreted that audits seem to be accepted as a preventive tool. Do the audits used in your organization provide numerical values, that can be interpreted as SPIs?
		1	*Yes*
		2	*No*

73 % of the surveyed organizations stated that they utilize a monitoring system for preventive safety activities. Considering the accidents and incidents during the last 365 days in your organization, were there any unforeseeable (black swan) events? (q_260572 - Typ 111)

v_124	v_7_1	int	73 % of the surveyed organizations stated that thy utilize a monitoring system for preventive safety activities. Considering the accidents and incidents during the last 365 days in your organization, were there any unforeseeable (black swan) events?
		1	*There were none*
		2	*1-2 events*
		3	*3-5 events*
		4	*6-10 events*
		5	*10 or more events*

45% of the organizations stated that there are areas in their operations which can't be measured by SPIs. What do you consider as not measurable by using SPIs? (q_260573 - Typ 142)

v_125	v_8_1	blob	45% of the organizations stated that there are areas in their operations which can't be measured by SPIs. What do you consider as not measurable by using SPIs?

callenge in def. / comparable (Hypothesis 3 a, b) (PGID 189843)

Although 82% do not use external indicators for comparison, the survey also showed that there is a demand for exchange of data, especially for operation's statistical data and safety management contribution and productivity as well as an exchange of risk exposure data. Which source of data do you consider as reliable? (q_260575 - Typ 121)

v_129	v_9_1	int	EASA
		0	*not quoted*
		1	*quoted*
v_130	v_9_2	int	the national aviation authority

		0	*not quoted*
		1	*quoted*
v_131	v_9_3	int	associations (e.g. EHA or EHAC)
		0	*not quoted*
		1	*quoted*
v_132	v_9_4	int	data from other organizations
		0	*not quoted*
		1	*quoted*
v_134	v_9_5	int	Others, like:
		0	*not quoted*
		1	*quoted*
v_128	v_9_6	var-char	Others, like:
v_133	v_9_7	int	none of them
		0	*not quoted*
		1	*quoted*

Each organization participating in the study has stated advantages in the exchange of data. Do you agree on sharing data from the following categories? (q_260576 - Typ 121)

v_140	v_10_1	int	Operation`s statistical data (e.g. mean time to failure, number of reports per year, serious incident rate, etc.)
		0	*not quoted*
		1	*quoted*
v_141	v_10_2	int	safety culture (e.g. commitment, justness, information, awareness, etc.)
		0	*not quoted*
		1	*quoted*
v_142	v_10_3	int	exposure to risk data (accepted missions in a certain risk profile, etc.)
		0	*not quoted*
		1	*quoted*
v_143	v_10_4	int	human resource data (staff turnover, postholder changes, etc.)
		0	*not quoted*
		1	*quoted*
v_144	v_10_5	int	contribution to safety management (over time)
		0	*not quoted*
		1	*quoted*
v_145	v_10_6	int	productivity of safety management (over time)
		0	*not quoted*
		1	*quoted*
v_146	v_10_7	int	effectiveness of safety management activity (over time)
		0	*not quoted*
		1	*quoted*
v_147	v_10_8	int	resilience indicators (e.g. number of staff with special competences)
		0	*not quoted*
		1	*quoted*
v_153	v_10_9	int	None of them
		0	*not quoted*
		1	*quoted*

Are there any restrictions in your organization concerning safety data exchange? (q_260577 - Typ 111)

v_148	v_11_1	int	Are there any restrictions in your organization concerning safety data exchange?
		1	*Yes, for example:*
		2	*No*
v_149	v_11_2	varchar	Yes, for example:

How do you rate the possibility to actually share information from 1-very unlikely to 5-very likely? (q_260578 - Typ 111)

v_150	v_12_1	int	How do you rate the possibility to actually share information from 1-very unlikely to 5-very likely?
		1	*1 - very unlikely*
		2	*2*
		3	*3*
		4	*4*
		5	*5 - very likely*

Admin (PGID 189767)

What is your position in the organization you work for? (q_260579 - Typ 141)

v_151	v_13_1	varchar	What is your position in the organization you work for?

Are there any additional comments you want us to consider for the follow-up examination? Is there anything you would like to address? (q_258097 - Typ 142)

v_83	v_14_1	blob	Are there any further comments, anything you would like to address?

Results of the second round of the survey

Refer to chapter 0 for a description of the survey results as second round of the survey is structured analogue to the first round.

The number of employees of the participating organizations range from less than 8 up to several hundred. This also implies different efforts in the safety management. How many full-time equivalent employees are currently engaged in your safety management? (q_260563 - Typ 111)	82% of the respondents of the first round stated that they consider the importance of the application of SPIs as high or very high. But only 73% actually use SPIs in their operation. Does your organization consider changing the amount of usage of SPIs in the future? (q_260564 - Typ 111)		
v_1_1	v_2_1	v_2_2	v_2_3
4	1	adherence to schedule, safety issues	-99
3	2	-99	-99
2	1	deviations from SOP´s, failure of barriers (Mitigations), Pilot induced Trigger in the Usage Monitoring System	-99
2	1	The current SPIs will be evaluated for validity	-99
1	2	-99	-99
2	2	-99	-99
1	2	-99	-99
1	3	-99	Maintenance, Planning, Staff (quality and safety)
6	2	-99	-99

In your opinion, what are the main reasons for using or not using SPIs? (q_260565 - Typ 142)
v_3_1
The complete acquisition and evaluation of the required data in order to create meaningful SPIs is very time-consuming.
We need SPIs for evaluating our SMS, (monitoring, measuring) and implementing safety performance targets and alert levels.
lack of time…
Monitor the low risk incidents that may escalate to serious incidents/accidents and applying risk controlling measures in advance
SPIs allow us to track our safety performance and also identify developing issues.
The limited resources in the safety management.
Effort
According to the accountable manager, the work of quality control and safety staff is unnecessary. Lack of communication. Nominated persons disagree. There is no professional standardization among nominated person.
SMS Department is understaffed in many cases.

Following the ICAOs approach on progress in aviation safety, a disaggregation of certain factors helps with the allocation in development stages of a safety system. Does your organization track any of the following influences using SPIs? (q_260566 - Typ 121)						
technical (technical and technological failures)	human factors (including man/machine interface errors)	organizational (systemic and culture impacts)	total system (failures due to the complexity of the system e.g.	Others, for example:	Others, for example:	none
v_4_1	v_4_2	v_4_3	v_4_4	v_4_5	v_4_6	v_4_7
1	1	1	1	0	-99	0
1	1	1	0	0	-99	0
0	0	0	0	0	-99	1
1	1	0	0	1	The most likely causes for an accident	0
1	1	0	0	0	-99	0
0	0	0	0	0	-99	1
0	0	0	0	0	-99	1
1	1	1	1	1	External - bird strike, laser strike, weather, another organization...	0
0	1	0	0	0	-99	0

SPIs can either be reactive and therefore based on historical data (like incidents in the company), or pro-active trying to disclose situations with higher or evolving risk. How many of the SPIs used in your safety management can be considered as proactive? (q_260567 - Typ 911)	Most organizations of the survey conduct between two to five audits per month, which could be interpreted that audits seem to be accepted as a preventive tool. Do the audits used in your organization provide numerical values that can be interpreted as SPIs? (q_260571 - Typ 111)	73 % of the surveyed organizations stated that they utilize a monitoring system for preventive safety activities. Considering the accidents and incidents during the last 365 days in your organization, were there any unforeseeable (black swan) events? (q_260572 - Typ 111)
v_5_1	v_6_1	v_7_1
44	2	2
24	1	2
0	1	2
83	2	2
50	1	2
0	2	1
26	2	1
71	1	2
33	2	2

45% of the organizations stated that there are areas in their operations which can't be measured by SPIs. What do you consider as not measurable by using SPIs? (q_260573 - Typ 142)	Although 82% do not use external indicators for comparison, the survey also showed that there is a demand for exchange of data, especially for operations statistical data and safety management contribution and productivity as well as an exchange of risk exposure data. Which source of data do you consider as reliable? (q_260575 - Typ 121)						
	EASA	the national aviation authority	associations (e.g. EHA or EHAC)	data from other organizations	Others, like:	Others, like:	none of them
v_8_1	v_9_1	v_9_2	v_9_3	v_9_4	v_9_5	v_9_6	v_9_7
The number of unreported events.	1	1	0	1	0	-99	0
Human factors	1	1	1	1	1	NTSB-BEA-LBA-CAA-	0
none	0	0	0	1	0	-99	0
Single, rare incidents	1	1	1	0	0	-99	0
Not sure.	1	1	1	1	0	-99	0
-	1	1	1	1	0	-99	0
-	1	0	0	1	0	-99	0
If the regulation provides clear, specific data and staff send reports.	1	1	1	0	1	ICAO	0
Fatigue related incidents	0	0	0	1	0	-99	0

Each organization participating in the study has stated advantages in the exchange of data. Do you agree on sharing data from the following categories? (q_260576 - Typ 121)								
Operation`s statistical data (e.g. mean time to failure, num-ber	safety culture (e.g. com-mitment, justness,	exposure to risk data (accepted missions in a certain risk	human resource data (staff turnover, postholder	contribution to safety man-agement (over time)	productivity of safety man-agement (over time)	effectiveness of safety man-agement activi-ty (over	resilience indicators (e.g. number of staff with special	None of them
v_10_1	v_10_2	v_10_3	v_10_4	v_10_5	v_10_6	v_10_7	v_10_8	v_10_9
1	1	1	0	1	1	1	0	0
1	1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	0
1	0	1	0	0	0	0	0	0
1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	0

Are there any restrictions in your organization concerning safety data exchange? (q_260577 - Typ 111)		How do you rate the possibility to actually share information from? (q_260578 - Typ 111)	What is your position in the organization you work for? (q_260579 - Typ 141)
Are there any restrictions in your organization concerning safety	Yes, for example:		
v_11_1	v_11_2	v_12_1	v_13_1
1	Muss im konkreten Fall geklärt werden	4	Safety Manager
2	-99	4	Safety and Compliance Manager Deputy
1	keine Herausgabe von Daten; alle Daten könnten von den Mitbewerbern für ihre Zwecke (z.B. Ausschreibungen) genutzt werden	1	Safety Manager
1	Personal data	4	Safety manager
2	-99	5	Line Pilot and Safety Manager
1	Exchange with other organizations.	3	Flight Safety Officer
2	-99	3	SM
1	anonymization of data	4	Compliance monitoring and safety manager
2	-99	5	Director Safety & Compliance

Are there any additional comments which you want us to consider for the follow-up examination? Is there anything you would like to address? (q_258097 - Typ 142)
v_14_1
-99
-99
no
-99
The exchange of data is really important particularly for small operators.
-
-
-99
-99