Monograph:

Gleirscher, Mario orcid.org/0000-0002-9445-6863 and Nyokabi, Anne (Submitted: 2019) System Safety Practice: An Interrogation of Practitioners about Their Activities, Challenges, and Views with a Focus on the European Region. Working Paper. arXiv . (Submitted)

System Safety Practice: An Interrogation of Practitioners about Their Activities, Challenges, and Views with a Focus on the European Region[*]

Preprint, compiled August 6, 2019

Mario Gleirscher[1] **and Anne Nyokabi**[2,3]

[1]Department of Computer Science, University of York, United Kingdom, Deramore Lane, Heslington, York YO10 5GH
[2]Process Industries and Drives Division, Siemens AG, Germany, Schuhstraße 60, 91052 Erlangen
[3]Formerly: Institut für Informatik, Technische Universität München, Germany, Boltzmannstr. 3, 85748 Garching bei München

Abstract

Context: System safety is the discipline for assessing the safety of socio-technical and software-intensive systems. While system safety has been a vital area of research for many decades, its practices are empirically not well studied. Beyond anecdotal evidence—case reports, interviews, forums, blogs—and insular surveys, we are missing large cross-disciplinary investigations that promote research validation and knowledge transfer. **Objective:** We explore means of work that safety practitioners rely on, factors influencing their performance, and their perception of their role. We examine observations from previous collaborations with industry and from the literature. **Methods:** We build a construct of system safety practice, collect data for this construct using an on-line survey, summarise and interpret the data, and investigate hypotheses based on the previous observations. **Results:** We present the responses of 124 practitioners in safety-critical system and software projects. Our respondents generally agree with statements such as: • safety decision making mainly depends on expert opinion and project memory, • safety is occasionally a cost-benefit question, • current safety standards reach their limits when applied in high automation domains, • assured reliability does not imply assured safety. Additionally, we contribute a research design directing towards explanatory studies of safety practice. **Conclusions:** We observe that empirical research of system safety practice requires more attention to mitigate the risk of undesirable mismatches between the state of the art and the state of the practice. This situation offers many new research opportunities.

***K*eywords** Safety-critical system · embedded software · empirical study · state of the practice · survey · questionnaire

1 Introduction

System safety practice (safety practice for short) is a remarkably diverse field spanning many disciplines involved in the system life cycle, influenced by heterogeneous criticality-driven *safety cultures* [1, 2, 3] across various application domains, geographical regions, and regulatory authorities.

Researchers have surveyed and investigated practised approaches to *accident prevention*, for example, in the chemical plant and nuclear power plant sectors [2] and in the construction industries [3]. However, our literature search has not uncovered a single officially published empirical investigation (i.e., a case or field study, a controlled field experiment, a survey of practitioners) of the *effectiveness of practised approaches* to prevent or reduce software and (control) systems' contributions to hazards.

In the following, we highlight the motivations for our study, describe observations from previous research, outline our research objective, and summarise the contributions of this work.

1.1 Motivation and Problem Statement

From previous discussions and research collaborations, we collected a variety of observations and developed several conjectures.

From exploratory content analysis of more than 200 selected *question and answer posts* on several safety practitioners' (SP) *on-line channels* between 2012 and 2016 and one expert interview [4], we observe that the members of the discussion groups

1. discuss various issues with the application of standards, calculation of failure rates, correct planning of safety tests, and completeness of hazard analyses;
2. are missing a standardised way of integrating safety with security activities;
3. are concerned about the adequacy of methods, a lack of safety education, and the misunderstanding of their role.

From exploratory content analysis of more than 370 *case reports* (i.e., on incidents and accidents) from the aviation (212), automotive (73), and railway (60) domains, published between 2000 and 2016, and from 7 *semi-structured interviews* with SPs from these domains [5], we observe that

4. human errors and specification errors were more often reported as accident root causes than software implementation errors—this is consistent with the findings in [6, 7, 8];
5. no IT security problems were reported as root causes—a case for further investigation in the light of the variety of known vehicle security threats [9, 10];

[1]* This work is partly supported by the Deutsche Forschungsgemeinschaft (DFG) under the Grant no. 381212925.
Reference Format: Gleirscher, M., & Nyokabi, A.. System Safety Practice: An Interrogation of Practitioners about Their Activities, Challenges, and Views with a Focus on the European Region. Unpublished working paper (August 6, 2019). Department of Computer Science, University of York, UK. arXiv: 1812.08452 [cs.SE]

*correspondence: mario.gleirscher@york.ac.uk

6. reports in general, and comparatively often in the automotive domain, were non-informative of subtle accident root causes (i.e., causes lying outside the possibilities, budgets, or obligations of accident analysts and investigators);
7. more than 30 of the selected reports suggest that accidental complexity [11]—particularly, missing or mistaken maintenance, refactoring, evolution, or migration—negatively affects system safety;
8. interviewees report issues of unclear separation of systems and software engineering tasks [12];
9. interviewees state that available methods are appropriate in their domains but can easily get insufficient for future high-automation applications.

These observations remind of the variety of computer-related risks [13] and of worries about the state of the practice and education in safety [14] and software [15] engineering (SE). However, these observations are anecdotal evidence and, while some being obvious, yet with limited justification for being valid, reliable, or general.

1.2 Research Objectives

Motivated by the observations and conjectures summarised in Section 1.1, we aim to explore safety practice and, where possible, seek for further evidence. Inspired by the research agenda in [16], our exploration starts from three questions addressing all of the issues 1 to 9:

1. Which means are SPs familiar with and which do they use? How effective are those means?
2. What are the SPs' challenges and expectations?
3. How do SPs view their profession and contribution in the life cycle?

1.3 Contributions

We present results from a cross-sectional self-administered online survey among safety practitioners. We compare their experiences, opinions, and their self-perception with previous observations and conjectures about safety practice (Section 1.1).

From these conjectures, we derive several hypotheses (Section 4.4) on safety practice and its practitioners. We analyse and test these hypotheses based on the gathered data and interpret the results (Section 5.1) with respect to previous experience and existing evidence. We justify the questionnaire and the derived hypotheses using existing research.

Furthermore, we respond to the request from Alexander et al. [17] and Rae et al. [18] for applying improved methodology in empirical research of safety practice, as well as the desire of a stronger involvement of SPs in research evaluation such as stated by Martins and Gorschek [16].

Finally, we contribute a research design (Section 3) for follow-up assessments with the potential of application to other SE domains (see, e.g. [19]).

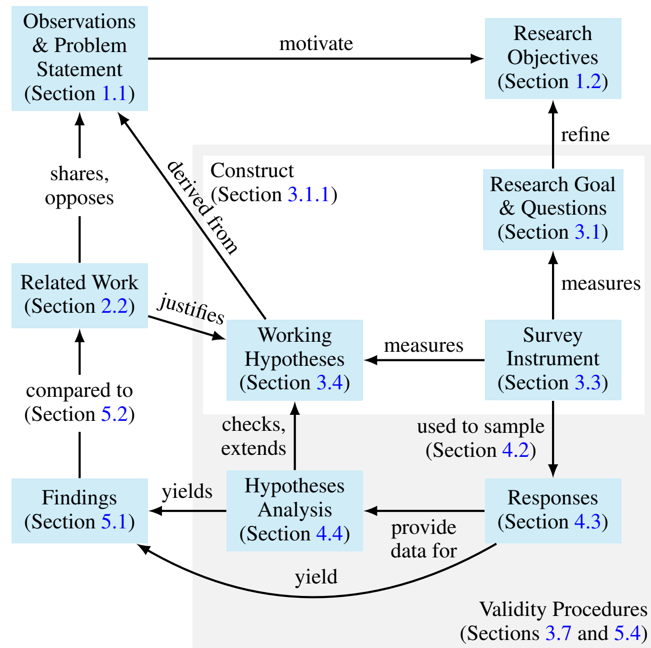


Figure 1: Overview of the research method for this article

1.4 Overview

Figure 1 provides an overview of the research procedures for this article. After discussing terminology and related work in Section 2 and describing our research method in Section 3, we present our results in Section 4. Particularly, we describe our sample in Section 4.2 and summarise the results of all valid responses in Section 4.3. Section 4.4 highlights the results of several hypotheses tests. Our discussion follows in Section 5, with the interpretation of our test results in Section 5.1 and the examination of threats to the validity of our study in Section 5.4. We summarise our findings in Section 6. Appendix A contains a detailed summary of the response data.

2 BACKGROUND

We introduce important terms as well as related work we will revisit in our discussions below.

2.1 Terminology and Definitions

The *life cycle* of an *engineered system* typically refers to the phases of design, implementation, release, maintenance, operation, and disposal. *Dependability* then encompasses the handling of reliability, availability, maintainability, and safety in the life cycle, for example, by improving fault-tolerance [20, 21].

In this work, we focus on the discipline of *system safety* [22],[2] including but not restricted to *functional safety*. System safety is situated in the context of safety of machinery,[3] process

[2]From software, electrical, electronics, control, and systems engineering.
[3]From mechanical engineering.

safety,[4] structural safety,[5] or occupational health and safety. These disciplines have in common the identification, assessment, and management of operational risk. This procedure includes the prevention or handling of *undesired events* at any execution stage of a system (e.g. hazards or safety risks, incidents, and accidents) and of any type (e.g. human error, software faults, and system failures). In addition, *security of information technology* (IT security or *security* for short) is the discipline of protecting computer-based systems and data against malicious attacks and unauthorised access.

Then, *safety practice* denotes the practical aspects of system safety in both industrial settings and applied research. Based on this, we consider a *safety practitioner* as a person who supports or performs *safety decision making*, particularly, by identifying *hazards* and assessing their causes and consequences, the design of hazard countermeasures (also known as hazard controls), the assurance of safety, or by performing research and consultancy for these *safety activities*. Importantly, there are many *means*—that is, best practices, methods, techniques, and standards—to apply in these activities.

2.2 Related Work

As indicated in Section 1, we found only few cross-disciplinary exploratory inquiries of safety practice and its practitioners. The following studies demonstrate the importance of empirical methods (interviews and related survey methods such as focus groups and questionnaires) in further examining safety practice.

Dwyer [23], for multiple disciplines, and Knight [12], for software engineering, characterise safety practice from their experience, forecasting the ongoing trend of increased automation, the increasingly critical interplay between the involved engineering domains, and the corresponding challenges for future safety research.

Adequacy of Means of Work in Safety Practice Safety-critical systems are subjected to *automation* (i.e., the use of qualified and verified tool chains) for their development, testing, and overall assurance. Graaf et al. [24] and Kasurinen et al. [25] investigate obstacles to the adoption of new methods and tools in embedded system RE, architecture design, and software testing. Our study explores this direction within safety practice.

Hatcliff et al. [26] summarise particular challenges in the certification of software-dependent systems and suggest improvements, stressing the concept of "designed-in safety/security." These works inspired and underpin our hypotheses but are different from our survey approach to examining safety practice and its practitioners.

Chen et al. [27] report on the challenges and best practices of using assurance cases. Our questionnaire about safety practice complements their study by exploring methods, training, and interaction, backed by a larger set of data points.

Ceccarelli and Silva [28] propose a framework for maturity assessment based on compliance checking during and after the introduction of new safety standards (e.g. DO-178B) into organisations that engineer safety-critical systems. In our study, we are asking SPs whether safety standards known and used by our respondents, actually improve an organisation's safety practice.

McDermid and Rae [14] report on their cross-domain insights into the practice of engineering safety-critical systems, discussing the question: "How did systems get so safe despite inadequate and inadequately applied techniques?" Not presuming that modern systems are acceptably safe, we interrogate SPs about their means of work.

Wang and Wagner [29] investigate decision processes in safety activities. For complex and highly critical systems such processes are usually committee- or group-driven to reduce organisational single points of failure. The authors examine whether such decision making is prone to a number of pitfalls known as "groupthink" and studied in group psychology. While our study does not apply a psychology-based construct, our observation of the strong reliance on expert opinion strengthens their conclusions.

Process Factors influencing Safety Practice Requirements engineering (RE) and, particularly, requirements specification, are critical points of failure in every safety-critical system project [7, 8, 6]. Examining research on the communication and validation of safety requirements in industrial projects, Martins and Gorschek [16] observe a lack of evidence for the usefulness and usability of recent safety research. We want to contrast their finding with how practitioners perceive the adequacy of their means of work.

Nair et al. [30] present results from a survey of 52 SPs on how they *manage the variety of safety evidence* for critical computer-based systems. Good evidence management implies to tackle traceability for change impact analysis (CIA), that is, the analysis of how changes of safety-critical artefacts (e.g. specifications, issue databases, designs) are propagated and whether these changes have negative safety impact. Borg et al. [31] report on 14 interviews with SPs about their CIA activities, finding that SPs have difficulties in understanding the motivation of CIA, are overwhelmed by the information they have to process when conducting CIA, and struggle with trusting and updating former CIAs. From a cross-sectional survey of 97 practitioners, De la Vara et al. [32] observe insufficient CIA tool support. Our study examines such means of work from a more general viewpoint.

Huber et al. [33] interviewed 8 automotive safety and security practitioners to find out how an integration of safety and security activities can avoid undesired incidents. The authors observe significant deficits of this integration, particularly, deficits in the traceability of the impact of security-related system changes on system safety.

In the Sections 3.3.1 and 3.4.1, we further relate these works with our study. In Table 8 in Section 5.2, we compare their findings with our results.

3 Survey Planning

This section describes the survey design (Section 3.1), the survey instrument (Section 3.3), the working hypotheses (Sec-

[4] From automation and plant engineering.
[5] From construction or civil engineering.

Table 1: Classification criteria for characterising the population and for sample assessment. **Legend:** MC... multiple-choice, (N)ominal or (O)rdinal scale

Classification Criterion	Scale
Educational Background	N / MC
Application Domains	N / MC
Level of Experience	O / duration in years
Familiarity with Standards	N / MC
Familiarity with Methods	N / MC
Geographical Regions	Open / MC
Native Languages	N / MC
Working Languages	N / MC
Safety-related Roles	N / MC

Table 2: Constituents of safety practice and practitioner's expectations and challenges. **Legend:** (N)ominal or (O)rdinal scale, (T)ruth values as nominal scale, * ... half-open or open.

Construct	Scales
Constituents of Safety Practice	
Safety Process (activities, roles, and practitioners)	N / e.g. decisions, hazard identification, resources
Factors (constraints and issues)	T / e.g. lack of resources, high schedule pressure
Means (conventional techniques; formal methods; tools; norms; skills; knowledge sources)	N* / e.g. FMEA, ISO26262, FMEA expertise, expert opinions
Application domains (current, new, complex)	N* / e.g. systems based on adaptive control, machine learning
Expectations & Challenges in Safety Practice (as perceived by SPs)	
Performance of safety activities	O / high ... low performance
Adequacy of means	O / high ... low adequacy
Collaboration between safety and security engineers	O / effective ... ineffective collaboration
Value of knowledge sources to SPs	O / high ... low, per class of methods or standards
Adaptation and improvement of SPs' skills	O / high ... low self-improvement/adaptation
Notion, perception, and priority of safety activities	N*
Contribution of SPs to system life cycle	O / high ... low contribution

tion 3.4), the procedure for data collection (Section 3.5) and analysis (Section 3.6), and instrument evaluation (Section 3.7). We follow the guidelines in [34] for planning and conducting the survey and [35, 36, 37] for the reporting.

3.1 Research Goal and Questions

The observations and conjectures summarised in Section 1.1 pertain to potentially critical issues in *safety practice*. This cross-sectional survey aims at resuming these issues. The *objective* of our exploration is

> to investigate *safety practice and its practitioners* and to examine observations we made during our preliminary research.

For this, we explore three *research questions*:

RQ1 Which *means* do SPs typically rely on in their activities? How helpful are those means to them?

RQ2 Which typical *process factors* have influence on SPs' decisions and performance?

RQ3 How do SPs perceive and understand their role in the *process* or life cycle?

3.1.1 Construct

For this objective and these research questions, we introduce the construct *safety practice and its practitioners (SPP)*. This construct incorporates SPs' processes, tasks, roles, methods, tools, and infrastructures and, by interrogating them via a questionnaire, their experience with and opinions about safety practice. SPP is divided into three sub-constructs: Classification of SPs, Constituents of safety practice, and Expectations & challenges in safety practice. The construct is visualised in Figure 2.

The criteria for classification in Table 1, the break-down in Table 2, and following research design are based on our research experience in system safety, on collaborations with industry, on expert interviews [38, 5, 4], and on previous work [39]. Relevant data from these unpublished works is either presented here or officially archived in [40]. The derivation of SPP follows the grounded theory approach [41].

Below, we use prefixes for referencing content items: RQ for research questions, h for working hypotheses, q for questions in the questionnaire, and F for findings. References have the shape $\langle X\rangle\langle Label\rangle[-\langle o\rangle]$ where $X \in \{RQ, h, q, F\}$ and o can refer to an answer option in the questionnaire.

3.2 Survey Participants and Population

Safety practitioners are our direct study subjects, our target group. A *safety practitioner* is a person whose professional activities as a practitioner or researcher in industry or academia are tightly related to the engineering of safety-critical systems. Table 1 lists criteria we use to characterise and identify members of the *population* of SPs. *Safety practice*, as described in Section 2.1, is our indirect study object. SPs participating in our study are also called *study or survey participants* or *respondents*.

3.3 Data Collection Instrument: On-line Questionnaire

Table 4 provides details on the (q)uestions. For traceability to the construct developed according to grounded theory and for concise presentation, we consolidated the original questions, taking care of maintaining their original meanings. For verification of this transformation, the original questionnaire [39] is archived in [40].

3.3.1 Motivations underlying the Questions

In the following, we establish links between the questions and the research summarised in Section 2.2.

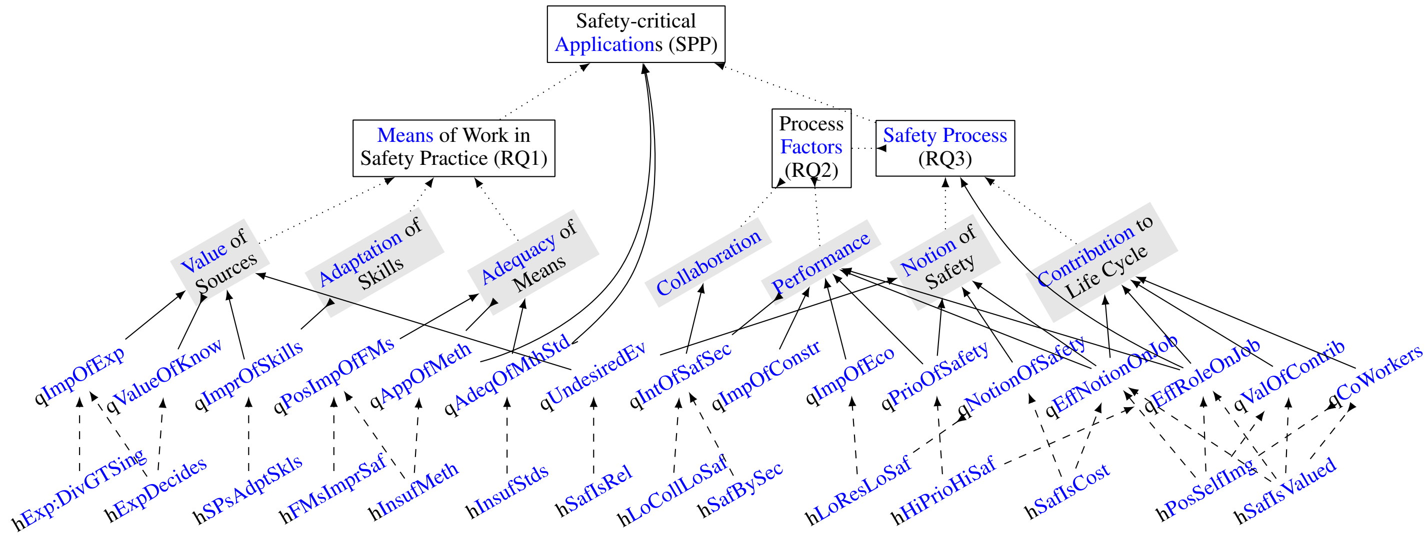


Figure 2: Research design for the construct "safety practice and its practitioners" (SPP, Section 3.1.1). The base (h)ypotheses layer *is backed by data of* the (q)uestionnaire layer (dashed edges). The latter layer contains questions *providing data about* (solid edges) expectations and challenges (boxes in grey). Expectations and challenges are *formulated over* (dotted edges) the Constituents of safety practice (framed boxes). For the sake of brevity, classification criteria (Table 1) are omitted.

Table 3: Scales used in the questionnaire

Type	Values
value	very high (*vh*), (*h*)igh, (*m*)edium, (*l*)ow, very low (*vl*)
agreement	strongly agree (*sa*), (*a*)gree, neither agree nor disagree (*nand*), (*d*)isagree, strongly disagree (*sd*)
impact	(*h*)igh, (*m*)edium, (*l*)ow, (*n*)o impact
adequacy	very adequate (*va*), (*a*)dequate, slightly adequate (*sa*), not adequate (*na*)
frequency	often, rarely/occasionally, never; or all, many, few, none
choice	single/multiple: (*ch*)ecked, (*un*)checked; or yes, no

qAdeqOfMthStd Bloomfield and Bishop [42] contrast prescriptive regulation with goal-based regulation, reviewing practice, highlighting potential benefits of safety cases along with the challenge of gaining sufficient confidence. Taking a more general position, question qAdeqOfMthStd is about *norms adequacy*.

For maturity measurement, Ceccarelli and Silva [28] work with a construct similar to the one in Table 2. By asking question qAdeqOfMthStd, we cover practitioners' opinions independent of a specific norm.

The questions about *adequacy of means* (particularly, qAdeqOfMthStd, qAppOfMeth, qPosImpOfFMs), aim at the re-examination of known challenges as, for example, discussed by Kasurinen et al. [25] and Graaf et al. [24].

The answer categories for question qAdeqOfMthStd are based on industry sectors with a relatively high pace of innovation and/or new, complex, but not yet well-understood system applications (e.g. self-driving cars).

qValueOfKnow Lethbridge et al. [43] observe that test and quality documentation is the most likely maintained kind of documentation. With question qValueOfKnow, we want to find out about the role of project documentation in safety decision making.

Rae and Alexander [44] examine how confidence in safety expert judgements (e.g. individual versus group judgements) is justified and leads to actual validity of the conclusions the further stages of the safety life cycle are based on. The authors argue that expert risk assessments exhibit low effectiveness in measuring risk as an objective quantity and propose "risk assessment as a means of describing, rather than quantifying risk." Their analysis extends the background of qValueOfKnow.

qIntOfSafSec and qPrioOfSafety While Chen et al. [27] focus on the aspect of *training and collaboration in safety assurance*, our study crosses these aspects with the questions qIntOfSafSec and qPrioOfSafety about interaction in and efficiency of safety activities.

The questions qIntOfSafSec, qValOfContrib, and qCoWorkers address the integration of safety activities with the life cycle, similar to Bjarnason et al. [45] on the alignment of RE and verification and validation.

In contrast to tool support for optimal auditing as investigated by Dodd and Habli [46], our questions (i.e., qEffRoleOnJob, qValOfContrib, qCoWorkers, and qImpOfExp) help to solicit personal views of SPs as external auditors and consultants.

qImpOfConstr, qImpOfEco, and qNotionOfSafety As summarised in Section 1.1, we presume negative consequences of "accidental complexity" [11] on system safety. Lim et al. [47] examine the perception of technical debt, highlighting the inevitable trade-off between software quality and business value. In an unfortunate case, an acceptance of technical debt can lead to an acceptance of low software quality, and

Table 4: Transcription and summary of selected questions from the questionnaire. **Legend:** Nominal, (O)rdinal, (L)IKERT-type scale, (T)ruth values as nominal scale, MC. . . multiple-choice, * . . . half-open or open. Figures 12 to 23 show details on the options; Sec./Fig. serves the navigation.

Question	**Scale** (see Table 3)	**Sec.**	**Fig.**	**N**
qValueOfKnow: Of how much *value* are specific *knowledge sources* for safety decision making?	L* / value per source	4.3.1	12	97
qImpOfConstr: To which extent do specific *process constraints and issues* negatively *impact safety activities*?	O* / impact per factor	4.3.2	13	93
qImpOfEco: How often do *economic factors* have a strong influence on the *handling of hazards*?	O / frequency	4.3.3	–	93
qAdeqOfMthStd: Regarding a specific *application domain*, how *adequate* are applicable safety *standards and methods* in ensuring safety?	O / adequacy per domain	4.3.4	14	102
qAppOfMeth: The *application of conventional techniques* (e.g. FMEA and FTA) has become too difficult for *complex applications of recent technologies*.	L / agreement	4.3.5	15	97
qPosImpOfFMs: Estimate the *positive impact* of formal methods on safety activities and system safety.	O / impact	4.3.6	16	58
qImprOfSkills: Specify your level of agreement with 4 statements about *factors improving a SP's skills*.	L / agreement per statement	4.3.7	17	96
qIntOfSafSec: Specify your level of agreement with 10 statements about the *interaction of safety and security* activities.	L / agreement per statement	4.3.8	18	95
qNotionOfSafety: How is safety *viewed* in your field of practice?	Nominal* / MC	4.3.9	19	95
qPrioOfSafety: Specify your level of agreement with 4 statements about *factors increasing the efficiency in safety activities*.	L / agreement per statement	4.3.10	20	97
qEffRoleOnJob: Is your *job affected* by any predominant definition of your *role*? In either case, we request for comment.	T* / comment	4.3.11	–	91
qEffNotionOnJob: Is your *job affected* by any predominant *view of safety*? In either case, we request for comment.	T* / comment	4.3.12	–	95
qUndesiredEv: Specify your level of agreement with 5 statements about *safety activities*.	L / agreement per statement	4.3.13	21	97
qValOfContrib: Of how much *value* is your role as a practitioner or researcher in safety-critical system developments?	L / value	4.3.14	22a	95
qCoWorkers: How much *value* do non-safety co-workers attribute to the *role of a safety practitioner*?	L / value	4.3.15	22b	95
qImpOfExp: Specify your level of agreement with 2 statements about the *role of experience in safety activities*.	L / agreement per statement	4.3.16	23	96

for some systems, to an acceptance of accidental complexity. Whenever this reasoning applies to a safety-critical system, we should ask whether this system is taken in by an unacceptable trade-off between safety and business value? Asking the questions qImpOfConstr, qImpOfEco, and qNotionOfSafety, we inversely probe the demand for investigations of the safety impact of technical debt.

Based on the SPP construct, we interrogate SPs about supportive factors (qValueOfKnow, qPrioOfSafety) and obstacles (qImpOfConstr) in safety decision making.

3.3.2 Notes on the Questionnaire

Some questions in Table 4 are *half-open*, that is, we allow respondents to extend the list of given answer options by using an extra text field. Some questions are *open*, that is, we only provide a single text field. Most demographic questions are half-open multiple-choice (MC) questions, that is, they have a text field "Other".

The scales used for encoding the answers in the column "Scale" are described in Table 3. We treat value and agreement as a 5-level LIKERT-type scale. Value, impact, adequacy, and frequency scales are equipped with a "do not know (*dnk*)" option. Together with "neither agree nor disagree (*nand*)" answers, participants are given two ways to stay indecisive. This way, we try to reduce bias by forced responses. From comparative analysis, we conclude that it is safe to **discard** *dnk*-answers and missing answers from our analyses.

We expect survey participants to spend 20–30 minutes on the questionnaire. Although we do not collect personal data, they can leave us their email address if they want to receive our results.

3.4 Working Hypotheses

We derive working hypotheses from our previous observations (Section 1.1). Table 5 contains two types of *(h)ypotheses* we want to analyse and test with the data we collect from the survey participants. The *base hypotheses* include observations or assumptions based on our own previous work experience or made by other researchers (Section 2.2). Additionally, we elaborate *comparative hypotheses* during exploratory analysis [48] of the responses.

Some hypotheses in Table 5 are measured by a single question (see, e.g. hExp:DivGTSing and qImpOfExp). We do not collect data for each individual construct referred to in such hypothesis/question pairs.

Figure 2 summarises the survey design presented in Sections 3.1 to 3.4 by showing important interrelationships between the base hypotheses, the questions of the questionnaire, and the parts of the SPP construct.

3.4.1 Motivations underlying the Hypotheses

Here, we justify the derivation of our hypotheses through links to other research (Section 2.2).

hExpDecides: SPs' activities mainly depend on expert opinion and experience from similar projects It is well-known that experts are fallible (see, e.g. recent investigations in [49, 44, 29]) and, thus, relying on experts in organisational (and engineering) decision making can contribute to critical single points of failures in such organisations. Moreover, it is well-known that reusing (e.g. cloning) repositories from finished projects in similar new projects bears many risks of errors in reuse or update of these data. Our previous interviews suggest that these knowledge sources are used in safety practice.

hLoResLoSaf: A lack of resources has a negative impact on the performance of safety activities The observations in Section 1.1 motivate the collection of evidence on whether or not *a lack of resources might have a negative impact on safety activities*. "Negative impact" refers to deferred safety decisions, hindered hazard identification and implementation of hazard controls, or limited SPs' abilities to fill their role. The conjecture that *budgets constrain safety activities* is further inspired by "the willingness to accept some technical risks to achieve business goals" as concluded by Lim et al. [47, p. 26].

hInsufStds: Safety activities for highly-automated applications lack support of appropriate standards and methods The belief that *safety practice is missing adequate standards and methods* has been discussed by Cant [50] and Knauss [51]. Questions about the *appropriateness of methods and standards* have also been raised by McDermid and Rae [14]. The idea behind hInsufStds is to understand the situation of SPs in new, not yet matured industry sectors. SPs would have the opportunity to adapt their skills and to gain further expertise (hSPsAdptSkls).

hSPsAdptSkls: SPs improve their skills towards new applications, e.g. by studying recent results in safety research Hatcliff et al. [26] observe that "industry's capability to verify and validate these systems has not kept up" (we inquire *willingness to improve skills* with hSPsAdptSkls) and that "the gap between practice and capability is increasing" because of *more integrated and more complex software technologies*. In contrast to the compliance framework presented by Ceccarelli and Silva [28], Hatcliff et al. highlight that showing compliance with existing norms cannot guarantee safety. Our study touches *norms adequacy* with hInsufStds.

hInsufMeth: Conventional methods (e.g. FMEA, FTA) are challenging to apply to complex modern applications The observation that *conventional methods have become inadequate* is broached by Knight [12, 52]. Likewise, McDermid and Rae [14] and Hatcliff et al. [26] underpin hInsufStds and hInsufMeth, though not the long-standing [53] and frequent expectation that *formal methods (FM) have a positive impact on safety practice* (hFMsImprSaf).

hFMsImprSaf: The use of formal methods has a positive impact on the performance of safety activities The *efficacy of FMs in practice* has been an only moderately researched subject for many years, investigated, for example, by Barroca and McDermid [54] and Woodcock et al. [55]. One intention underlying hFMsImprSaf is to determine whether we have to further examine FM effectiveness to cross-validate reported experiences (e.g. [56]).

hSafBySec: For current applications, the assurance of safety also depends strongly on the assurance of security Safety-critical applications of networked or connected (software) systems have lately revived the question of how safety and IT security influence each other (see, e.g. [33])? Along these lines, the justification of hSafBySec is based on manifold *anecdotal evidence* (see, e.g. [57]) that security problems can cause safety violations and, possibly, vice versa.

hSafIsCost: Safety is more seen as a cost-increasing rather than a cost-saving part in many application domains How are the practical achievements and implications of *system safety and the effort spent therefor* related? How relevant are such utilitarian and controversial questions to SPs and their organisations? Touching this subject, hSafIsCost is formulated in the context of "total cost of safety," that is, the cost of accident prevention and accident consequences borne by organisations that engineer and operate safety-critical systems. hSafIsCost's truth might contribute negatively to the role of SPs in an (engineering) organisation.

hLoCollLoSaf: A lack of collaboration of safety and security engineers has a negative impact on safety activities According to Conway [58], the structure of an engineered system converges towards the (communication) structure of its engineering organisation. For example, in a safety-critical distributed embedded system (e.g. avionics, process automation, and automotive architectures), team collaboration would determine the architectural decomposition and direct communication links in the architecture. However, team collaboration not necessarily implies keeping track of the impact of critical changes across all critical relationships. It is known (e.g. [33]) that critical relationships in a complex architecture are far from obvious. Such relationships are sometimes only indirectly perceived as an undesired *emergent property*. Hence, we ask SPs about the collaborations among so-called "property engineers," e.g. safety and security engineers (qIntOfSafSec).

hSafIsRel: SPs understand safety as a special case of reliability Leveson [22] stresses an observed misconception about system safety, namely that the responsibility to make systems *safe enough* is reduced to the responsibility to make their *critical parts just reliable enough*. Her claim stimulates the question to which extent SPs are solely driven by reliability concerns and which negative implications this might have. Moreover, hSafIsRel is also motivated by examinations [59] of how findings from previous accidents can be included in safety arguments.

hAdapt:AutoGTAero: SPs using automotive standards agree more than SPs using aerospace standards that skill adaptation is required and takes place From several pre-

Table 5: Overview of hypotheses (used as H_1 in the tests). **Legend**: See Section 3.3. The quantification ranges a–j refer to the answer options of the questions associated with the hypotheses, see Figures 12 to 23. The original questionnaire is archived in [40].

Hypothesis	**Supported if** ... (AC, Section 3.6.2)
Base Hypotheses	
hExpDecides: SPs' activities *mainly depend on* (d) expert opinion and (g) experience from similar projects.	$\forall o \in \{d, g\}$: $\mathsf{med}(q\mathit{ValueOfKnow}^o) \in \{h, vh\} \wedge o$ among 3 highest valued (of 7) knowledge sources $\wedge$ $\mathsf{med}(q\mathit{ImpOfExp}^a) \in \{a, sa\}$
hLoResLoSaf: There is a lack of resources *that has a negative impact on* the performance of safety activities.	$\forall o \in \{a, d\}$: $\mathsf{med}(q\mathit{ImpOfConstr}^o) \in \{m, h\} \wedge \mathsf{med}(q\mathit{ImpOfEco}) \in \{\mathit{often}\} \wedge q\mathit{NotionOfSafety}^c \leq 30$
hInsufStds: Safety activities for highly-automated applications *lack support of* appropriate standards and methods.	For ≥ 5 out of 7 domains o: $\mathsf{med}(q\mathit{AdeqOfMthStd}^o) \in \{sa, na\}$
hInsufMeth: Conventional methods (e.g. FMEA, FTA) *are challenging to apply* to complex modern applications.	$\mathsf{med}(q\mathit{AppOfMeth}) \in \{a, sa\} \wedge q\mathit{PosImpOfFMs}_{m+h} > 25\%$
hFMsImprSaf: The use of formal methods *has a positive impact on* the performance of safety activities.	$\mathsf{med}(q\mathit{PosImpOfFMs}) \in \{m, h\}$
hSPsAdptSkls: SPs improve their skills *towards* new applications, e.g. *by* studying recent results in safety research.	$\forall o \in \{a, b\}$: $\mathsf{med}(q\mathit{ImprOfSkills}^o) \in \{a, sa\}$
hSafBySec: For current applications, the assurance of safety *also depends strongly on* the assurance of security.	$\forall o \in \{a, c, e, f\}$: $\mathsf{med}(q\mathit{IntOfSafSec}^o) \in \{a, sa\}$
hSafIsCost: Safety *is more seen as* a cost-increasing *rather than* a cost-saving part in many application domains.	$q\mathit{NotionOfSafety}^a_{ch} > 60\% \wedge \forall o \in \{b, e\}$: $q\mathit{NotionOfSafety}^o_{ch} < 40\%$
hLoCollLoSaf: A lack of collaboration of safety and security engineers *has a negative impact on* safety activities.	$\forall o \in \{h, i, j\}$: $\mathsf{med}(q\mathit{IntOfSafSec}^o) \in \{a, sa\}$
hHiPrioHiSaf: Prioritisation of safety in management decisions enables SPs to *perform their tasks more efficiently*.	$\forall o \in \{a, b\}$: $\mathsf{med}(q\mathit{PrioOfSafety}^o) \in \{a, sa\}$
hSafIsRel: SPs understand safety as *a special case of* reliability.	$\forall o \in \{a, e\}$: $\mathsf{med}(q\mathit{UndesiredEv}^o) \in \{a, sa\} \wedge \forall o \in \{b, c\}$: $\mathsf{med}(q\mathit{UndesiredEv}^o) \in \{d, sd\}$
hSafIsValued: SPs believe that their non-safety co-workers *attribute high value* to SPs' contributions.	$\mathsf{med}(q\mathit{CoWorkers}) \in \{h, vh\} \wedge \mathsf{med}(q\mathit{ValOfContrib}) \in \{m, h, vh\}$
hPosSelfImg: SPs perceive their *contribution as highly valuable*.	$\mathsf{med}(q\mathit{ValOfContrib}) \in \{h, vh\} \wedge \mathsf{med}(q\mathit{CoWorkers}) \in \{m, h, vh\}$
Comparative Hypotheses	
hExp:DivGTSing: SPs with high diverse expertise *better perform in safety activities*) than SPs with low singular expertise.	$\forall o \in \{a, b\}$: $\mathsf{med}(q\mathit{ImpOfExp}^o) \in \{a, sa\} \wedge \mathsf{med}(q\mathit{ImpOfExp}^c) \in \{\mathit{nand}, a, sa\}$
hValue:SenLTJun: *Senior SPs* attribute lower value to their role in the system life-cycle than *junior SPs* (cf. hSafIsValued, hPosSelfImg).	One-sided **U** succeeds with $p < 0.05$
hAdapt:SenGTJun: *Senior SPs* agree more than *junior SPs* that skill adaptation (e.g. learning) is required and takes place (cf. hSPsAdptSkls).	One-sided **U** succeeds with $p < 0.05$
hAdapt:AutoGTAero: SPs using *automotive standards* agree more than SPs using *aerospace standards* that skill adaptation (e.g. learning) is required and takes place (cf. hSPsAdptSkls).	One-sided **U** succeeds with $p < 0.05$
hInsufMeth:EngDifSci: *Engineering-focused SPs* agree different from *research-focused SPs* with hInsufMeth.	Two-sided **U** succeeds with $p < 0.05$
hInsufMeth:AutoGTAero: SPs using *automotive standards* agree more than SPs using *aerospace standards* with hInsufMeth.	One-sided **U** succeeds with $p < 0.05$

vious discussions with SPs, we learned the view that system safety practice in the automotive domain is for several reasons less extensive than in other domains, such as aerospace. Hence, we assume that automotive SPs are more strongly involved in or aware of skill development in their domain than SPs in aerospace.

3.5 Data Collection Procedure: Sampling

To draw a diverse sample of safety practitioners, we

1. advertise our survey on safety-related on-line discussion channels,
2. invite practitioners and researchers in safety-related domains from our social networks, and
3. ask these people to disseminate information about our survey.

Our sampling procedure can best be described as a mixture of opportunity, volunteer, and cluster-based sampling. The cluster is formed by survey participants from several of these channels. We expect our sample to be stronger than non-randomised but, because of a lack of control of the sampling process, weaker than uniformly random.

Sample Representativeness To check how well our final sample *appropriately represents safety practice and its practitioners*, the questionnaire measures the classification criteria in Table 1. See [40] for the questions used for this.

3.6 Analysis Procedure

This section describes the analysis of the responses, the checking of the working hypotheses, and our tooling.

3.6.1 Analysis of Responses

We use instruments of descriptive statistics [60] such as median (med), mean (μ), variance (var), and frequency histograms to summarise the responses per question.

Half-Open and Open Questions We use answers from the text field "Other" to extend and revise the classifications imposed by the given answer options. See Section 4.2 for the results. Furthermore, we close some of the main questions using qualitative content analysis and coding [61]. For some half-open questions, we extend the statement lists and nominal scales accordingly. The results of this step are shown in Section 4.3 when discussing the questions in the Sections 4.3.1, 4.3.2, 4.3.9, 4.3.11 and 4.3.12.

3.6.2 Hypothesis Analysis and Statistical Tests

We use *non-statistical analysis* for all base hypotheses for which we directly[6] collect data (Table 5).

For most comparative hypotheses, we apply the MANN-WHITNEY **U** test [60] (**U** for short) to check for difference. We use **U** if the following assumptions hold:

- exactly one LIKERT-type or ordered-categorical dependent variable (DV),
- random division into two groups,
- group members are not paired,
- treatments via independent variables (IV) are already applied,
- group sizes may differ and be small (< 30),
- per-group distributions of the DV may be dissimilar and non-GAUSSIAN.

Let H be a hypothesis and α be the maximum chance of a Type I error, that is, incorrect rejection of the null hypothesis H_0. **U** tries to reject H_0 with a confidence of $1 - \alpha$. We require $p < \alpha$ for the Type I error p of *incorrectly distinguishing two groups of respondents* with respect to H. If **U** succeeds to reject H_0 then the **support of the desired alternative hypothesis** H_1 is increased. Failure of **U** in rejecting H_0 (i.e., $p \geq \alpha$) denies any conclusion on H from the given data set [62, p. 168]. The medium maturity and criticality of our hypotheses (for an exploratory study) and the medium accuracy of our data (from a survey method) justify $\alpha = 0.05$.

Acceptance Criteria (AC) The criteria in Table 5 describe the aggregation of the question scales in Table 3 to match the hypotheses. These criteria are built from symbols of the kind $q\langle id\rangle^{\langle answer\ option\rangle}_{\langle scale\ value(s)\rangle}$ referring to the questions in Table 4. We require med to be non-central to express a large supportive majority. Alternatively, percentage thresholds (e.g. > 25%) express the desired variance or shape of the distribution. In hypothesis tests, we mainly use classification criteria (Table 1) as IVs.

3.6.3 Tooling

We use Unipark[7] as a platform for implementing on-line surveys and for data collection (Section 3.5) and temporary storage. For statistical analysis and data visualisation (Section 3.6.2) we use GNU R[8] and Unipark. Content analysis and coding takes place in typical spreadsheet applications.

3.7 Validity Procedure after Survey Planning

In the following, we evaluate the face and content validity of our instrument, and the internal and construct validity of our study. Although, we did not perform an independent pilot study according to [36], we took several measures to assess the validity of our study.

3.7.1 Instrument Evaluation: Face and Content Validity

Both authors performed several internal walk-throughs to improve the survey design and the data analysis procedure. Along the lines of a *focus group*, we asked independent persons to complete the questionnaire and to provide feedback via an extra form field in the questionnaire and via email. This dry run took place between 13 and 27 June 2017. We gathered 7 responses, from 2 postgraduate research assistants with experience in the survey method, and with experience in safety-critical software,

[6] For example, hypothesis hInsufMeth refer to *Adequacy of means*. To keep our questionnaire lean, qAppOfMeth directly measures *agreement* for one *instance* of this hypothesis.

[7] See http://www.unipark.de.

[8] See https://www.r-project.org.

systems, and requirements engineering, 1 master student with industrial work experience in safety-critical systems engineering, 1 IT practitioner and English native speaker, 1 person with a health and safety background, 2 practising software engineers.

The feedback from the these respondents resulted in

- an extension and balancing of answer options,
- the alignment of answer scales throughout the whole questionnaire,
- improvement of the nomenclature (terms are now described on the questionnaire page they first appear).
- an extension of open answer fields, and
- linguistic improvements.

These steps helped us to improve questionnaire completeness, consistency, and comprehensibility and to reduce researcher bias. The full questionnaire is available in [40].

3.7.2 *Internal Validity of the Analysis Procedure*

Why would the procedure in Section 3 lead to reasonable and justified results?

U is applicable only if groups are independent with respect to the considered IV. Hence, when using MC-questions as IVs in comparisons, we ensure a partitioning of the sample by excluding data points shared between groups. For example, with qNotionOfSafety as IV, respondents who are in both groups "responses with choice (c)" and "responses with choice (a)" are omitted.

The 7 test data points allowed us to validate our tooling (e.g. R scripts, see Section 3.6.3). Test data points are not included in the final data set.

3.7.3 *Construct Validity*

Why would the construct (Section 3.1.1) appropriately represent the phenomenon to investigate?

Because of the exploratory nature [48] of our study, the sub-constructs and their scales in Table 2 represent the study object as reconstructed from our analyses. The working hypotheses and the questionnaire represent an *approximation and a selection* of what needs to be measured and tested if we were to investigate this study object (cf. Figure 2) in an explanatory study. For example, we assume that the 10 statements in Figure 18 for question qIntOfSafSec satisfactorily approximate the "interaction of safety and security activities" (i.e., construct Collaboration) and its criticality. Consequently, the scales in Table 2 serve as a reference to the internal validation of our study.

The low strength of hypotheses derivable from our exploratory construct limits the strength of our conclusions. For example, an accepted hFMsImprSaf (i.e., FMs have a positive impact) reflects very much the personal experience, perception, or opinion of our survey participants. Their view has to be distinguished from the question of actual FM effectiveness. To pursue such a question in future work, we suggest to refine our construct using the *technology acceptance model* [63].

Table 6: Safety-related channels we advertised our survey on (sorted alphabetically by category, full list in [39, pp. 92f])

Channel Type	Example/References
Facebook sites	E.g. Int. Society of SPs
General panels	SurveyCircle, www.surveycircle.com
LinkedIn groups	E.g. on ARP 4754, DO-178, ISO 26262
Mailing lists	E.g. system safety (U Bielefeld,[9] formerly U York)
Newsletters	GI requirements engineering
Personal web-sites	E.g. profiles on Twitter, LinkedIn, Xing
ResearchGate	Q&A forums on www.researchgate.net
Xing groups	E.g. safety engineering
Other channels	E.g. board of certified safety professionals

Furthermore, an extended version of our inquiry of SPs about supportive factors and obstacles in safety decision making could include questions about "safety evidence traceability and management." Our questionnaire could therefore be extended by criteria examined in Nair et al. [30], De la Vara et al. [32] and Borg et al. [31].

The sub-construct Constituents resembles the framework for safety process maturity assessment as proposed by Ceccarelli and Silva [28]. By deriving our questionnaire from qualitative analysis of several discussions and interviews with practitioners, we follow an established research methodology (see, e.g. [64]). Overall, we believe our design is appropriate with respect to the expressive power of the working hypotheses and serves as a good starting point for successive explanatory studies.

3.7.4 *Reliability*

A check for test-retest reliability (e.g. changing attitudes of respondents) and alternate form reliability are out of scope of this exploratory study. Hence, we do not plan to ask respondents to answer the questionnaire more than once and we run only one variant of the questionnaire.

4 Survey Results

In this section, we characterise our sample (Section 4.2), summarise the responses (Section 4.3), and analyse our hypotheses (Section 4.4).

4.1 *Survey Execution: Sample Size and Response Rate*

For the collection of data from the participants, we

1. advertised our survey over the channels in Table 6 and
2. personally invited more than 20 persons.

The sampling period lasted *from 1 July 2017 til 25 September 2017*. In this period, we *repeated step 1 up to three times* to increase the number of participants. The Unipark tracking data shows that LinkedIn groups, ResearchGate, Twitter, and mailing lists were effective in soliciting respondents, however, it does not disclose which channels were most effective.

[9] See http://www.systemsafetylist.org.

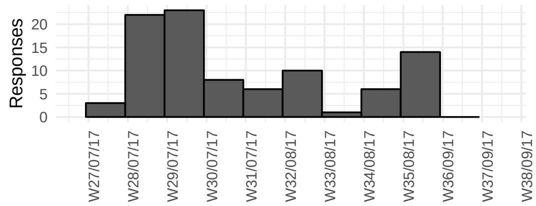


Figure 3: History of responses

After 565 views of the questionnaire, our final sample contains $N' = 124$ (partial) responses with $N = 93$ completed questionnaires and $N = 91$ (73%) complete[10] data points. Figure 3 depicts the distribution of responses over time. According to Unipark, respondents spent 20 minutes on average to provide complete data points, 50% spent within 14 and 24 minutes time.

Based on the member count of specific channels (e.g. for LinkedIn groups), the return rates of responses per channel range from 0.1 to 5%.

According to the classification criteria, we will reason about sub-groups of the sample with at least 15 data points. Appendix A provides the data summaries (i.e., number of answers per option) for all questions.

4.2 *Description of the Sample*

We describe our sample in the following and estimate the extent to which it represents (Section 3.5) the population of SPs. For each classification criterion in Table 1, we provide a chart or we name the up to 10 most frequently occurring answers, ordered by frequency. Percentages (%) right of the bars indicate the fraction of the 93 completed questionnaires, shown in parentheses the (N)umber of respondents who chose the corresponding option. Note that most of the classification questions allow MC answers.

Educational Background Figure 4 summarises the *educational background* of all respondents:

- *Computer scientists* include software engineers and computer engineers
- *Electrical and electronics engineers*
- *Safety scientists* include safety engineers, occupational safety practitioners, health and safety practitioners, human factors engineers, ergonomics engineers
- *Mechanical and aerospace engineers*
- *Systems engineers* include poly-technical systems engineers, information systems engineers, business technologists, engineering business administrators, engineering project managers
- *Physicists and mathematicians*
- *Other discipline* includes chemists, biochemists, civil engineers, language scientists

[10] Apart from two options of the classification question Safety-related Roles (66, 76) and the question qPosImpOfFMs (62), we had at least 91 up to 124 responses for each question.

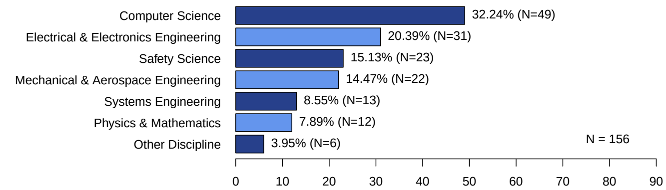


Figure 4: Educational Background (frequency, MC)

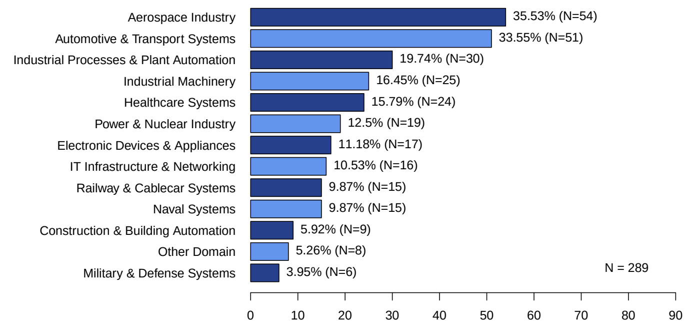


Figure 5: Application Domains (frequency, MC)

Application Domains Figure 5 summarises the *application domain* of all respondents where "aerospace" includes space telescopes; "industrial processes and plant automation" includes manufacturing, chemical processes, oil and gas, energy infrastructure, and small power plants; "railway systems" includes railway signalling; "construction and building automation" includes civil engineering applications; and "other domains" includes food safety, biological safety, research and development, and environment, health, and safety preparations.

Level of Experience Figure 6 indicates that our sample of SPs is moderately balanced across all experience levels.

Familiarity with Standards Figure 7 provides an overview of *safety-related* standards our respondents are familiar with (distinguished by generality or by application domain): Standards from aerospace (e.g. ARP 4761, DO-178, DO-254), generic standards (e.g. ISO 61508, DIN VDE 0801) automotive (e.g. ISO 26262), machinery (e.g. ISO 13849, 25199, DIN EN 62061, MRL 2006/42/EG), military (e.g. MIL-STD 882, UK Def Std 00-55), railway (e.g. CENELEC EN 50126, 50128, 51029, 62061), power plants (e.g. IEC 60880, 61513, 62138, 60987, 62340, IEC 800), and medical devices (e.g. IEC 80001, ISO 14971, AAMI/UL 2800). 14 participants were neither familiar with any of the given standards nor did they specify other standards.

Familiarity with Methods Figure 8 shows the familiarity of our respondents with prevalent *concepts of safety analysis* and the corresponding *classes of methods*, techniques, or notations:[11] For example, FMEA, FMECA, or FMEDA to assess *failure mode effects*; HazOp studies, ergonomic work analysis and intervention methodology to assess *hazard operability*; *STAMP-based methods* for hazard (STPA) and accident (CAST) analysis; FHA, FFA, PHA, or PHL to assess risk at a *functional or abstract design level*; common cause (CCA)

[11] Abbreviations are described in Table 9 in Appendix B.

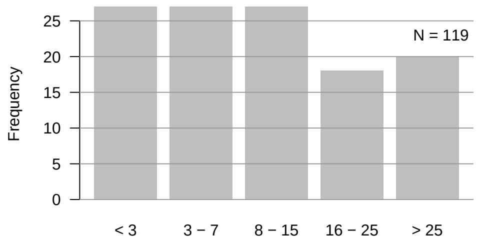


Figure 6: Level of Experience (time intervals in years)

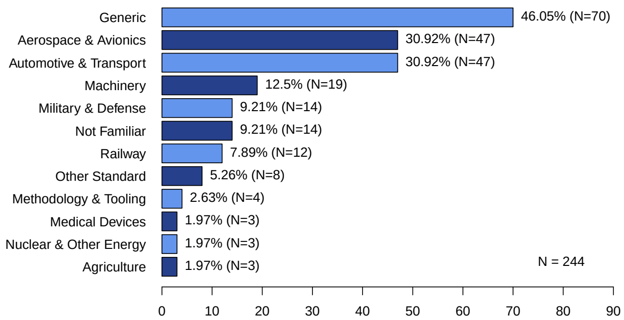


Figure 7: Familiarity with Standards and use by domain (frequency, MC)

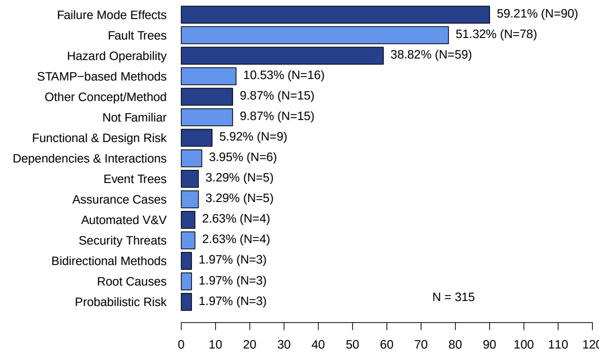


Figure 8: Familiarity with Methods and concepts (frequency, MC)

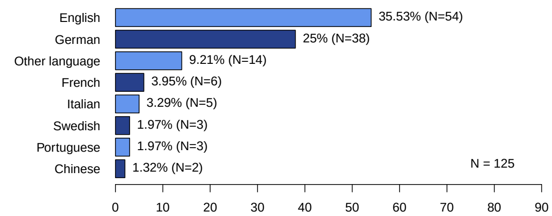


Figure 9: Native Languages and concepts (frequency, MC)

or common mode (CMA) analysis to include *dependencies and interactions*; fault injection and property checking as techniques of *automated* validation and verification (V&V); STRIDE or CORAS to assess and handle *security threats*; *bidirectional* methods such as Bowties or cause-consequence analysis; MARKOV chains for *probabilistic* risk analysis, and GSN and SACM to build *assurance cases*. For "Other", our participants mentioned a variety of approaches (no more than twice): 5S, 5W, CASS, coexistence analysis, FRAM, HazRAC, HEART, HRA, MTA, (O)SHA, SAR, SCRA, SHARD, SSHA, Poka Yoke, prognostic analysis, WBA, ZHA, ZSA.

Only 4 respondents state familiarity with methods to assess and handle *security threats*. 15 respondents *neither checked any of the given methods nor did they specify other methods* that are relevant in their safety activities.

Geographical Regions DE (24.3%), UK (16.4%), US (15.3%), AU (6.2%), FR (5.1%), IT (3.4%), CA (3.4%), CN (2.8%), and CH (2.8%).

Native Languages Figure 9 provides an overview of the *languages spoken* by the respondents.

Working Languages Figure 10 provides an overview of the *languages used at work* by the respondents.

Safety-related Roles In Figure 11, the term *practitioner* includes the profile of an engineer and a manager. Regarding engineering *disciplines* and domains, "safety practitioner" includes engineers or managers in system safety, functional safety, or in other safety domains as well as technology risk managers in general; "software practitioner" includes developers, architects, and tool developers; "systems practitioners" includes system analysts and system architects; "health & safety practitioner" includes occupational safety practitioners, human factors engineers, and ergonomists; and "V & V practitioner" includes test and assurance practitioners. For "Other", our respondents include a civil engineer, a project manager, a method engineer, and a maintainability engineer.

Regarding *responsibility profiles*, the category "Consultant / Assessor" includes independent evaluators, auditors, regulators, and inspectors dealing with safety certification.

4.3 Summary of Responses

In this section, we summarise the responses to the questions in Table 4.

Guide to the Figures The following text and figures complement each other. For LIKERT-type scales, we use centred diverging stacked bar charts as recommended by [65]. med denotes the median and "ex" indicates the number of excluded data points per answer option.

4.3.1 qValueOfKnow: Value of Knowledge Sources

Figure 12 shows that, among the knowledge sources we asked our participants to rate, *expert opinion*, *previous experience in safety-related projects*, and *case reports* represent the three highest valued knowledge sources used in safety activities and safety decision making. *Management recommendations* turn out to be the lowest valued knowledge source.

The following knowledge sources, or resources in more general, were additionally mentioned to be of *very high or high value*:

Four respondents referred to the concept of *adversarial thinking*, mentioning "creative mind", "imagination", "analysis capability," and "acceptance of human fallibility." Three respondents pointed to the concept of *domain expertise and expe-*

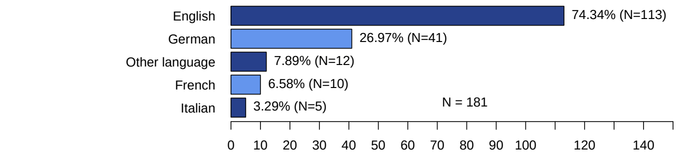


Figure 10: Working Languages and concepts (frequency, MC)

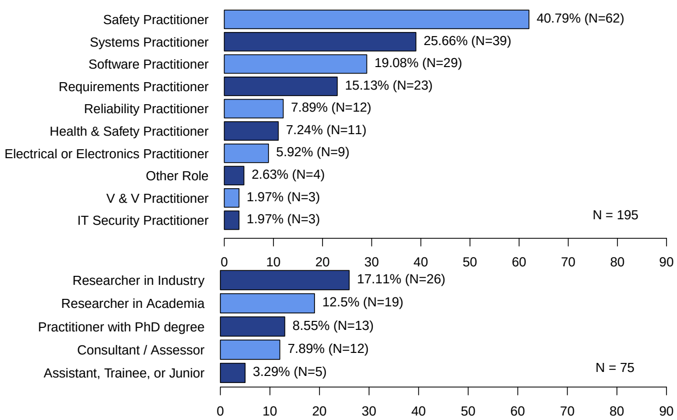


Figure 11: Safety-related Roles (frequency, MC) split into disciplines (top) and responsibility profiles (bottom)

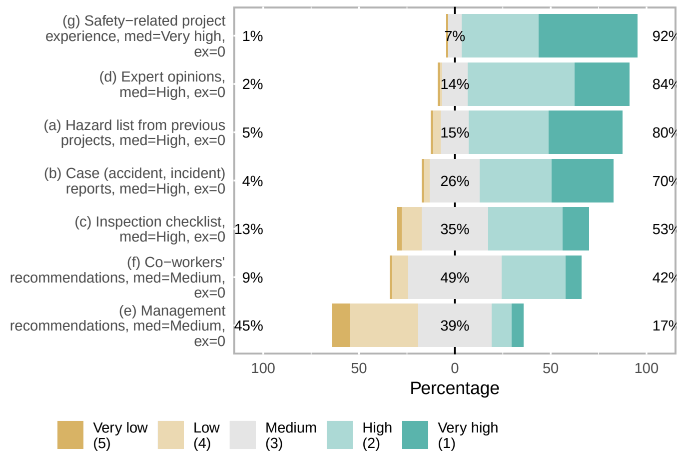


Figure 12: qValueOfKnow ($N = 97$): Value of knowledge sources – Of how much *value* are specific *knowledge sources* for safety decision making?

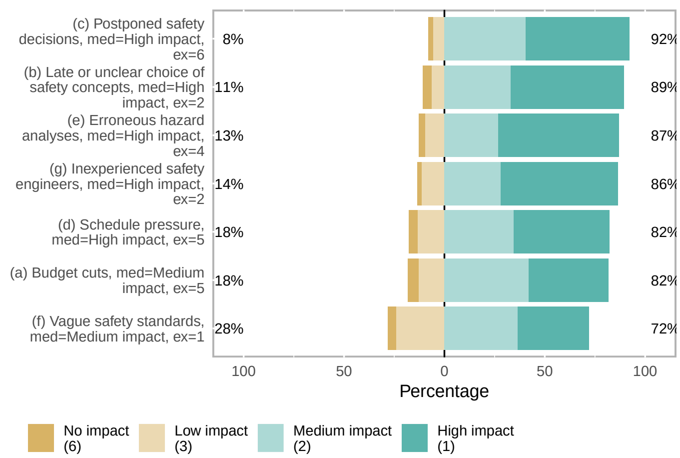


Figure 13: qImpOfConstr ($N = 100$): Negative impact on safety activities – To which extent do specific *process constraints and issues* negatively *impact safety activities*?

rience, mentioning "gut feel", "subject matter knowledge of the application," and "...real work and related problems in reference situations..." Furthermore, they mentioned *education*, *specification documents and tools* (e.g. "use of SPARK"), *independent assessment*, *in-service monitoring logs*, and *previously certified similar systems*.

4.3.2 *qImpOfConstr: Constraints on Safety Activities*

According to Figure 13, *inexperienced safety engineers (g)* and *erroneous hazard analyses (e)* gained the most ratings in the category "significant negative impact on safety activities." *Postponed safety decisions (c)* achieved the largest consensus. *Vague safety standards (f)* constitutes the bottom of this ranking but is still rated with medium or high negative impact by the majority of respondents.

The following factors (i.e., process constraints and issues) were additionally mentioned to have *high negative impact* on safety activities:

Eight respondents broach the issue of *missing management expertise and support*: "Lack of education of managers in need for safety" identifies one respondent from the oil and gas industry. Another one states that there is a "general perception that safety is only paper work" and perceives a "lack of safety knowledge within management." One practitioner was even pointing to a "lack of general safety culture."

Three participants criticise that *the degree of collaboration is too low*: They perceive a "lack of system level engineering experience" as well as "soloed working practices without a clear view of [an] integrated safety concept" and that the organisation is "minimising [the] involvement of safety process/engineers into [the] development process."

Regarding *incomplete or inadequate hazard lists*, respondents mention "unidentified hazard domains" and "imagined safety cases not based on real workers experience." Along with that, one practitioner mentions the issue of "poorly defined requirements": Such requirements, when coming from upstream, are known to have a negative effect on many downstream engineering activities. Conversely, inadequate hazard lists resulting from such activities can again have a negative impact on downstream sub-system requirements specification.

Regarding *compliance with norms*, one respondent was criticising the "transfer of concern from assessment to compliance," in other words, *compliance bias*. Two others are broaching the opposite phenomenon of *compliance ignorance*, mentioning "general ISO 26262 standard ignorance" and a "lack of understanding of regulatory framework."

Furthermore, according to another participant's experience there is "too much faith in testing" and "reluctance to use formal methods."

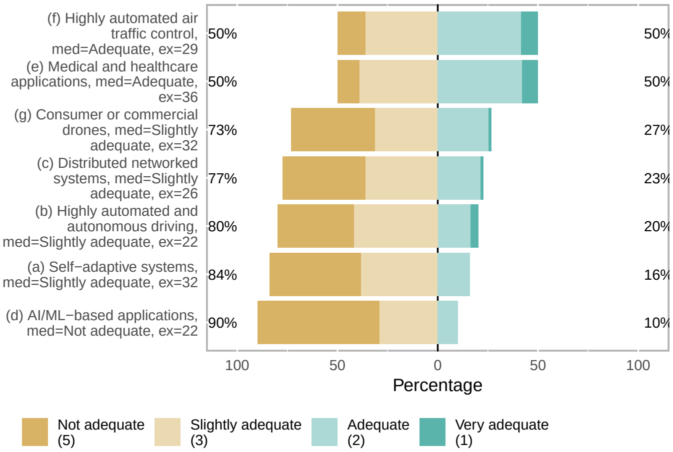


Figure 14: qAdeqOfMthStd (N = 102): Adequacy of methods and standards – Regarding a specific *application domain*, how *adequate* are applicable safety *standards and methods* in ensuring safety?

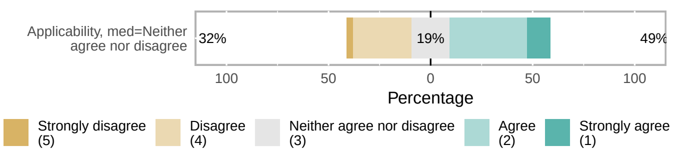


Figure 15: qAppOfMeth (N = 97): Applicability of methods – The *application of conventional techniques* (e.g. FMEA and FTA) has become too difficult for *complex applications of recent technologies*.

4.3.3 qImpOfEco: Influence of Economic Factors

More than a third (36%) of the survey participants share the view that economic factors *often strongly influence* the way how hazards are handled, about half of them (48%) think that such influence happens *rarely or occasionally* (median), and for 9% such influences are *not recognisable*.

4.3.4 qAdeqOfMthStd: Adequacy of Methods and Standards

According to Figure 14, traffic control (f) and medical and healthcare applications (e) are most often believed to be supported by *adequate* methods and standards. However, for all domains, at least 50% of the respondents think that the available means are only *slightly* or *not at all adequate* for safety assurance. This question exhibits a relatively large number of *dnk*-answers.

4.3.5 qAppOfMeth: Applicability of Methods

The *nand*-median in Figure 15 shows that there is no tendency or no clear consensus among respondents on whether or not conventional methods have become too difficult to apply in current applications.

4.3.6 qPosImpOfFMs: Positive Impact of Formal Methods

The median of "medium impact" in Figure 16 indicates a consensus among the participants on that the use of FMs might have a *positive* impact on the effectiveness of safety activities.

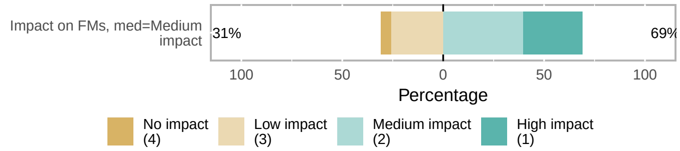


Figure 16: qPosImpOfFMs (N = 58): Positive impact of formal methods – Estimate the *positive impact* of formal methods on safety activities and system safety.

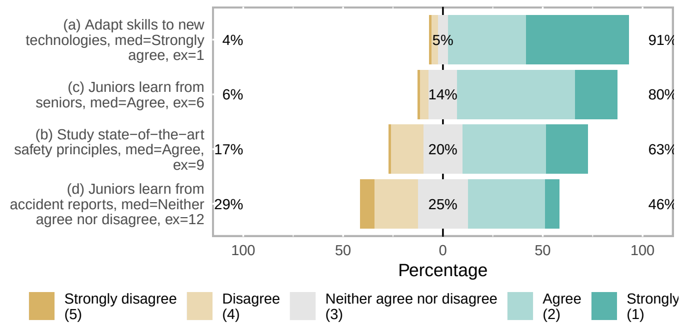


Figure 17: qImprOfSkills (N = 96): Improvement of skills – Specify your level of agreement with 4 statements about *factors improving a SP's skills.*

However, we only have a low number of responses resulting from missing answers and we excluded *dnk*-answers.

4.3.7 qImprOfSkills: Improvement of Skills

According to Figure 17, SPs agree moderately (39%) to strongly (54%) with the requirement to adapt their professional skills to new technologies. However, significantly less consensus was achieved among the respondents on whether junior SPs should learn from accident reports.

4.3.8 qIntOfSafSec: Interaction of Safety and Security

The high moderate and strong agreement in Figure 18 indicates that most of our participants perceive interactions between safety and security as critical.

SPs agree on that interaction between safety and security practitioners during requirements engineering and system assurance rarely occurs (g,f). Furthermore, agreement is achieved for the "negative influence of a lack of collaboration (between safety and security engineers)" (h,i) and for the "positive influence of such a collaboration" (j) on safety activities. However, we observe 7% of disagreement with the "requirement of ultimate IT security for safety."

No consensus is achieved regarding the *dependence of security on safety* (b,d). As opposed to that, respondents agree on the *dependence of safety on security* (a,c,e).

4.3.9 qNotionOfSafety: Notion of Safety

The answers in Figure 19 show that many participants seem to be reluctant to associating cost/benefit schemes with management decision making in system safety (a,b). Accordingly, many responses indicate that safety is treated as a cost-independent necessity (c). However, 51 (32%) responses were

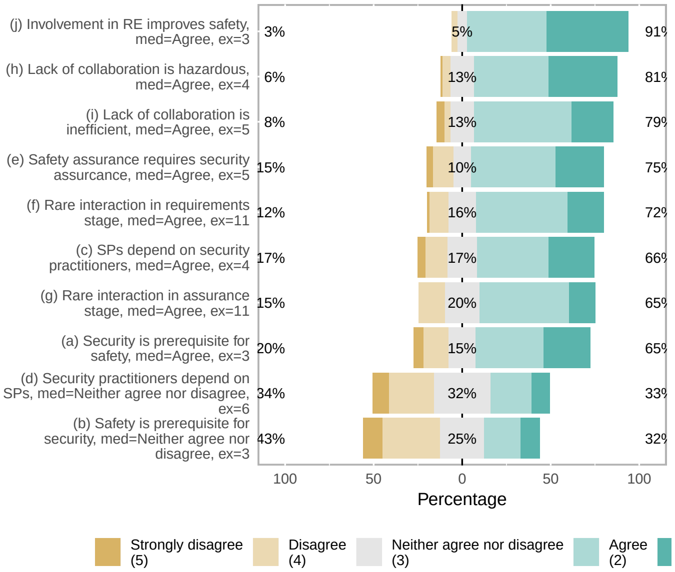


Figure 18: qIntOfSafSec (N = 95): Interaction of safety and security – Specify your level of agreement with 10 statements about the *interaction of safety and security* activities.

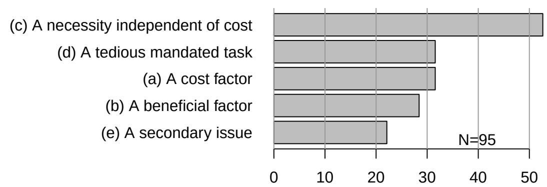


Figure 19: qNotionOfSafety (N = 95, MC): Frequency of safety notions – How is safety *viewed* in your field of practice? It is viewed as …

given to the view of safety as an "important, yet secondary, and tedious mandated issue" (d,e).

Beyond the five given answer options, the notions of safety additionally given by our respondents range from a "huge effort generating source", a "marketing gadget", a "high level product performance characteristic", a "regulation", a "general and common demand", a "must have" up to being "essential."

Importantly, two respondents add that *it depends* "on the manager or the engineer" or "on the stakeholder and on the safety professional." An ergonomist with 3 to 7 years of work experience says that "ergonomists usually are seen as added value to [the field] because we try to work to improve performance and health at the same time, safety is the natural outcome of this methodology."

4.3.10 qPrioOfSafety: Priority of Safety

Figure 20 indicates a consensus of the respondents for all given options (a–d). Particularly, increased priority of safety decisions (a) and defined safety processes (d) positively contribute to the efficiency of safety activities. The Sections 4.3.11 and 4.3.12 provide more details on the factors believed to increase the efficiency and effectiveness of safety activities **as well as dual factors** assumed to decrease the efficiency and

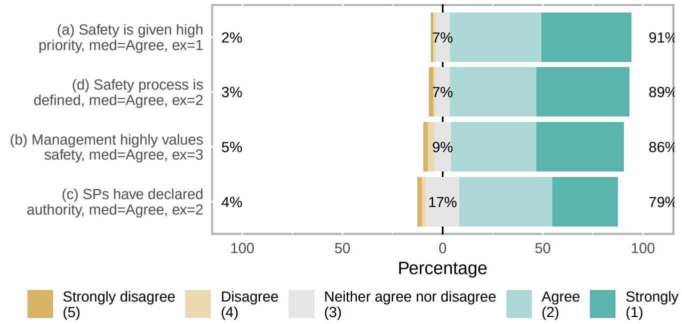


Figure 20: qPrioOfSafety (N = 97): Efficiency of safety activities – Specify your level of agreement with several statements about *factors increasing the efficiency in safety activities*.

effectiveness thereof. Along the way, comparatively many SPs (17%) do not offer any agreement on authority (c).

4.3.11 qEffRoleOnJob: Effect of Role Model on a SP's Job

We asked our respondents to comment on *whether and how their job is affected by the definition of their role*, if any, in their organisations and application domains.

Apart from 5 *dnk*-answers, we received 56 answers saying "yes" and, thus, stating that the role of a SP is *clearly defined*. These SPs perceive or expect the following *positive consequences* on their job (frequency given in parentheses, in descending order): Clear role definitions …

- have a general positive impact on a SP's activities (24),
- lead to clear responsibilities, authority, and escalation routes (13),
- allow good integration of safety activities into the surrounding system life cycle processes (6),
- can make the achievement of compliance easier (1), and
- let SPs maintain autonomy or independence to carry through their most critical activities (1).

However, our study participants report on the following *negative effects* on their job: Clear role definitions can …

- make engineers entirely push away safety-related responsibilities as a consequence of separating teams into safety and non-safety co-workers (2),
- lead to complex process definitions (1),
- get rather independent SPs exposed to company-wide resource and risk management (1), and
- impose a wrong focus or unnecessarily constrain a SP's tasks (1).

Moreover, 30 participants responded with a "no" and, hence, state that the role of a SP is *not clearly defined*. These SPs consider or expect the following *positive consequence* on their job: Unclear role definitions …

- may promote more freedom to act, for example, to develop and employ new and more effective safety approaches (3).

However, our respondents also perceive several *negative effects* on their job: Unclear role definitions ...

- can entail unclear or wrong responsibilities, limited authority and autonomy, and limited space for discretionary activity (9),
- promote unclear, one-sided, or late decision making, in the worst case, rushed processing of checklists (6),
- have a general negative impact on a SP's tasks (4),
- can lead to disintegrated conceptions of safety, separated communities with a lack of communication and coordination, promoting unnecessarily confined decisions (3),
- can decrease the appreciation of a SP's analysis capabilities (2), and
- increase the risk of unqualified personnel assuming the role of a SP (2).

4.3.12 *qEffNotionOnJob: Effect of Safety Notion on a SP's Job*

We asked our respondents to comment on *whether and how their job is affected by a predominant notion of safety* (qNotionOfSafety), if any, in their organisations and application domains.

Apart from 9 *dnk*-answers, we received 74 answers indicating a "yes" and, hence, stating that the notion of safety *has an effect* on their job: 10 respondents do not provide a specific comment. The others argue from several *notions of safety* they have perceived in their environments. Below, we provide answer frequencies and cite a few answers underpinning the summary statements.

Non-supportive Notions of Safety 24 participants describe their experiences with a *non-supportive, misunderstood, or underrated safety culture*. They report that ...

... *SPs have difficulties to argue their findings* (9): "Right now there is no ability to have the safety requirement override standard functional requirements." – "1. Our job always gets delayed and we are the last to get the inputs. 2. Non-safety engineers always try to justify or avoid the suggestions/findings. 3. It is difficult to sell safety culture to non-safety engineers/managers." – "I have to spend extra time explaining that safety is not about compliance or implementing controls." – "As for now safety has not the degree of importance to support testing views and arguments against system designers and management."

... *SPs suffer from late decision making* (5): "If I am not allowed to do my job early in the process (requirements stage), safety becomes more costly and I as a safety practitioner am viewed as a late check in the box to get through a program rather than an integral part of a design team."

... *SPs' activities have no lasting value* (1): "The safety practitioner is neither equipped, nor capable of making the decisions needed for a higher level of safety. Being mostly policemen, enforcers and rule designers, little if any of their contributions have any meaningful or lasting value."

Supportive Notions of Safety 20 respondents describe their experiences with or their view of a *supportive or highly-valued safety culture*. They report that ...

... *SPs' findings are important and heard* (6): "My job is important because safety is valued and considered necessary." – "Most people in my organisation understand the importance of safety. This is positive." – "There are not many people who practice safety, since it is a tedious job. So we are highly valued."

... *SPs are properly included in the process* (1): "Safety is fundamental to the work we do and is ingrained into our processes in such a way that its impossible to ignore. While it makes jobs harder with much more analysis and review processes and every stage of the product's development, we know its vital."

Other Notions of Safety 9 SPs describe an *ambivalent picture*, saying that it depends on individual projects whether their jobs are negatively or positively affected: "Safety at the last two places I worked is a check box activity at best. Other places I've worked it was started early in the pre-design phase. Starting early is more cost and schedule effective with a better end product."

5 SPs refer to a *regulation-driven* notion of safety: "Positively affected. In aerospace, safety is part of fundamental engineering principles, so the process is embedded in systems engineering and does not get left out."

From a *budget- or schedule-driven* perspective, respondents (4) observe that "the budget for tools and training is never enough" and that "resources, budget, support depend on the view/culture of safety."

Finally, 12 respondents claim, by saying "no", that the notion of safety does *not have any effect* on their jobs.

4.3.13 *qUndesiredEv: Role of Undesired Events for Safety*

Figure 21 shows a clearly disagreeing response on whether lack of failures reduces the need for carrying through safety activities (a). We have a more ambiguous agreement on whether *safety implies reliability* (e), that is, on whether having assured the safety of a system *usually includes* having also assured the reliability of a system. Moreover, known and reported accidents seem to be important for the argumentation of the need for safety (b,c). However, the agreement on whether a "lack of accidents weakens arguments for the need of safety" (d) varies more.

4.3.14 *qValOfContrib: Value of SPs' Contributions*

According to Figure 22a, the majority of respondents perceives their role in the system life cycle as highly valuable or better. The analysis and comments in the Sections 4.3.11 and 4.3.12 provide a more differentiated picture of this answer.

4.3.15 *qCoWorkers: Viewing SPs' Co-workers*

Figure 22b suggests that the respondents vary strongly in evaluating *their contributions to the system life-cycle* when trying to imagine their non-safety co-workers appreciation.

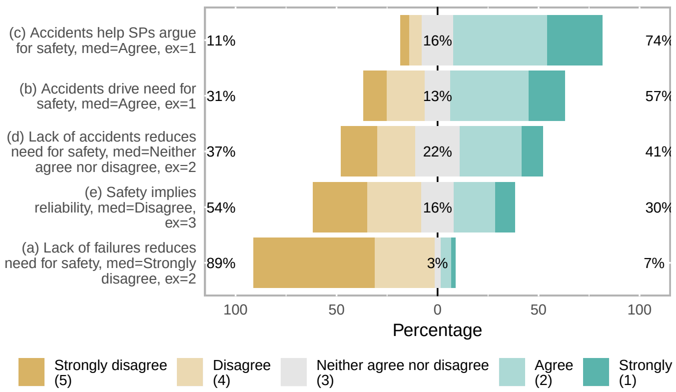
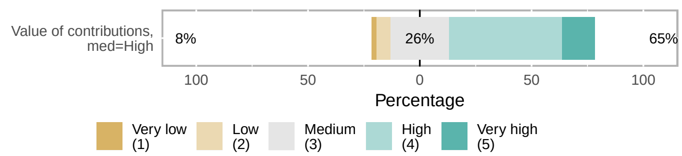
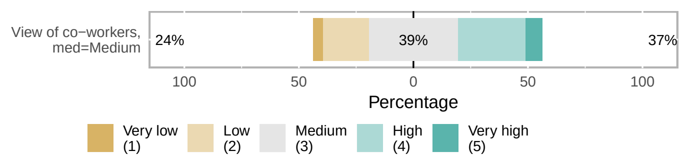


Figure 21: qUndesiredEv ($N = 97$): Role of undesired events (i.e., failures, incidents, and accidents) for safety – Rate your level of agreement with 5 statements about *safety activities*.



(a) qValOfContrib ($N = 95$): Value of SPs' contributions – Of how much *value* is your role as a practitioner or researcher in safety-critical system developments?



(b) qCoWorkers ($N = 95$): Viewing SPs' co-workers – How much *value* do non-safety co-workers attribute to the *role of a safety practitioner*?

Figure 22: Self-perception of SPs' role

4.3.16 *qImpOfExp: Influence of Experience*

From the responses, Figure 23 shows that experience in safety activities is believed to be positively associated with *improved hazard handling* (a,b), particularly, experience from similar previous projects (b). Adversarial thinking (c) receives the least agreement.

4.4 *Hypothesis Analysis and Test Results*

Table 7 presents the test results for all hypotheses listed in Table 5 and based on the summary in Section 4.3. Motivations for the acceptance criteria given in Table 5 are provided in Section 3.4.1. In summary, we were not able to find significant differences for the *pairs of groups (IVs) we compared* with respect to several DVs.

5 DISCUSSION

We interpret the responses (Section 5.1), draw a relationship to existing evidence (Section 5.2), and assess the validity of our study (Section 5.4). From these discussions, we derive our conclusions in Section 6.

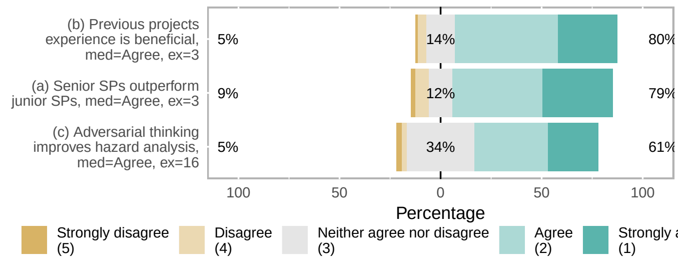


Figure 23: qImpOfExp ($N = 96$): Role of experience – Specify your level of agreement with 3 statements about the *role of experience and adversarial thinking in safety activities*.

5.1 *Interpretation of the Results and Findings*

The following discussion takes into account the hypothesis analysis and test results summarised in Table 7. Details about the hypotheses and the questions referred to in the text are given in the Tables 4 and 5.

5.1.1 *Findings for RQ1: Means of Work in Safety Practice*

Hypothesis hExpDecides is supported This should not be surprising as it mirrors a rather typical situation in many engineering disciplines and projects. However, relying too much on knowledge of experts can form a *single point of failure of an organisation*. Moreover, relying too much on experience from similar projects can go along with *wrongly transferring former conclusions (i.e., project memory) and not updating them* correspondingly.

Finding 1 *The responses suggest that safety strongly depends on expert opinion and project memory.*

Hypothesis hInsufStds is supported With regard to the given application domains (qAdeqOfMthStd, Section 4.3.4), the result for hInsufStds is negative: Our responses indicate that inadequate methods or standards constitute a real issue in current high-automation safety practice. However, from qImpOfConstr in Figure 13, we know SPs think that "vague safety standards" are problematic, though, least problematic of all inquired process constraints and issues. The 22 to 36 excluded *dnk*-answers might stem from the fact that most respondents can only make a statement for a small subset of the inquired application domains. We believe, the exclusion of these responses does therefore not influence our conclusion. Moreover, the observation of a lack of appropriate standards and certification guidelines is anecdotally confirmed by McDermid and Rae [14] and empirically in the automated vehicle testing domain by Knauss et al. [51, pp. 1878f].

Finding 2 *In the considered high automation domains, the adequacy of standards and methods for safety assurance is rated with low to very low.*

Hypothesis hInsufMeth is rejected Because of overlapping Means, the rejection of hInsufMeth contrasts the support of hInsufStds. We see a tendency towards our experience from collaborations with industry (Section 1.1) justifying hInsufMeth. We interpret the respondents' ambiguous

Table 7: Results of hypotheses analysis and tests. **Legend:** AC. . . acceptance criterion

Analysis of Base Hypotheses	**From the responses to . . .**	**. . . we conclude that . . .**
hExpDecides: Dependence on expert opinion	qValueOfKnow (Section 4.3.1) and qImpOfExp (Section 4.3.16)	our AC is **fulfilled**.
hLoResLoSaf: Resources govern performance of SPs	qImpOfConstr (Section 4.3.2) and qImpOfEco (Section 4.3.3)	the qImpOfEco-part of our AC is not fulfilled.
hInsufStds: Inadequate means in high-automation	qAdeqOfMthStd (Section 4.3.4)	our AC is **fulfilled**.
hInsufMeth: Low method adequacy	qAppOfMeth (Section 4.3.5) and qPosImpOfFMs (Section 4.3.6)	the qAppOfMeth-part of our AC is not fulfilled by the *nand*-median.
hFMsImprSaf: Positive impact of formal methods	qPosImpOfFMs (Section 4.3.6)	our AC is **fulfilled**.
hSPsAdptSkls: Necessity of skill adaptation	qImprOfSkills (Section 4.3.7)	our AC is **fulfilled**.
hSafBySec: Dependence on IT security	qIntOfSafSec (Section 4.3.8)	our AC is **fulfilled**.
hSafIsCost: Safety is a cost-benefit question	qNotionOfSafety (Section 4.3.9)	the qNotionOfSafety-a-part of our AC is not fulfilled.
hLoCollLoSaf: Benefit of safety-security interaction	qIntOfSafSec (Section 4.3.8)	our AC is **fulfilled**.
hHiPrioHiSaf: Benefit of safety-as-a-priority	qPrioOfSafety (Section 4.3.10)	our AC is **fulfilled**.
hSafIsRel: Safety is a special case of reliability	qUndesiredEv (Section 4.3.13)	none of the qUndesiredEv-parts of our AC are fulfilled.
hSafIsValued: High contribution to life cycle	qCoWorkers (Section 4.3.15) and qValOfContrib (Section 4.3.14)	the qCoWorkers-part of our AC is not fulfilled.
hPosSelfImg: High contribution (self-image)	qValOfContrib (Section 4.3.14) and qCoWorkers (Section 4.3.15)	our AC is **fulfilled**.
Test of Comparative Hypotheses	**From the comparison of . . .**	**. . . we conclude that . . .**
hExp:DivGTSing: Benefit of diverse expertise	senior SPs with junior SPs (from responses to qImpOfExp, Section 4.3.16)	our AC is **fulfilled**.
hValue:SenLTJun: Assoc. of expertise & value	senior SPs with junior SPs	with $p = 0.15$, our AC is not fulfilled.
hAdapt:SenGTJun: Assoc. of expertise & skill adaptation	senior SPs with junior SPs	with $p = 0.052$, our AC is **almost fulfilled**.
hAdapt:AutoGTAero: Assoc. of standards & skill adaptation	SPs using automotive standards with SPs using aerospace standards	with $p = 0.048$, our AC is **fulfilled**.
hInsufMeth:EngDifSci: Assoc. of profession & inadequate means	engineering-focused SPs with research-focused SPs	with $p = 0.22$, our AC is not fulfilled.
hInsufMeth:AutoGTAero: Assoc. of standards & inadequate means	SPs using automotive standards with SPs using aerospace standards	with $p = 0.54$, our AC is not fulfilled.

agreement with “available standards and methods have become too difficult” as “they can be challenging to apply.” However, from our data, we are unable to explain possible reasons for the adequacy or inadequacy of means.

Hypothesis hFMsImprSaf is supported The low number of valid responses to question qPosImpOfFMs limits the support of hFMsImprSaf. Both, the question qPosImpOfFMs and the notion of a formal method are abstract. Moreover, the classification questions do not provide enough information about our respondents’ experience with FMs.

> **Finding 3** *Among the informed respondents, formal methods are believed to be beneficial.*

The breadth of our exploration made it necessary to sacrifice the level of detail for certain questions, for example, qAppOfMeth, to keep the questionnaire short. Adequacy could be measured by a set of questions to facilitate a more fine-grained analysis. For example, asking for agreement in question qAppOfMeth can be substituted by asking for the *level of* Adequacy. Additionally, question qAppOfMeth could be asked for each technique and standard and analysed for sensitivity to industry-specific sub-groups of respondents. For example, the survey in [66] refines some of these ideas.

5.1.2 *Findings for RQ2: Impact of Process Factors*

Hypothesis hLoResLoSaf is rejected Four respondents to question qEffNotionOnJob report to have experienced a *lack of resources* for safety activities. This is consistent with the data checked for the AC of hLoResLoSaf. Although the responses suggest that the implication *lack of resources has negative impact on safety* might hold, the antecedent of this hypothesis is not broadly supported.

> **Finding 4** *The respondents suggest that resources occasionally but not typically govern SPs’ performance.*

However, by weakening hLoResLoSaf, we can acknowledge the “often” third of SPs showing a situation demanding for reaction in the community.

Hypothesis hSafIsCost is rejected We identify a weak positive association: Safety is most frequently viewed as *a cost-independent necessity* (qNotionOfSafety-c, hSafIsCost) and the median of qImpOfEco (hLoResLoSaf) lies at *economic factors rarely or occasionally influence safety.* So, for hSafIsCost, the many positive responses to the options (c,d,e) underpin the view of safety as a cost-independent factor in management decision making. We consider this to be positive but stress the need of a specific study of this finding.

Finding 5 *Our respondents suggest that safety is not typically a question of cost-benefit.*

Hypotheses hLoCollLoSaf is supported Our data supports hSafBySec, stating that safety assurance strongly depends on security assurance. Interestingly, for hLoCollLoSaf, SPs agree on both that . . .

Finding 6 *. . . a lack of collaboration downgrades the performance of safety activities (qIntOfSafSec-h,i),* ***and***

Finding 7 *. . . interaction between safety and security practitioners rarely occurs in requirements and assurance activities (qIntOfSafSec-f,g).*

Apart from desirable interactions at an organisational level, potential *dependence of security on safety* (qIntOfSafSec-b,d) is less obvious to our respondents than potential *dependence of safety on security* (qIntOfSafSec-a,c,e). While the latter is comparatively well known, the former is more difficult to grasp. Further investigations of this interaction based on practitioners' experience, such as [33], are necessary.

Finding 8 *Early collaboration of safety and security experts is viewed as essential.*

Hypothesis hExp:DivGTSing is supported Although the three propositions in Figure 23 seem obvious, we included them in the questionnaire to confirm that such assumptions are made by SPs (hExp:DivGTSing). The support of hExp:DivGTSing backs the support of hExpDecides.

Finding 9 *Our respondents see diverse expertise as essential for safety practitioners.*

Hypothesis hAdapt:SenGTJun versus hAdapt:AutoGTAero and hSPsAdptSkls The result for hAdapt:SenGTJun and hSPsAdptSkls is unsurprising because senior experts are expected to have seen paradigm shifts and technology changes more often than junior SPs. However, the small difference between both groups indicates that senior experts would avoid outdated skills as much as junior professionals would. The support of hAdapt:AutoGTAero suggests that automotive SPs see a stronger necessity of skill adaptation than aeronautic SPs. One reason for this development could be the currently raising demand for functional safety of increasingly complex driver assistance systems.

Finding 10 *Respondents from the automotive domain perceive adaptation of skills to new technologies to be more relevant for SPs than respondents from the aeronautics domain.*

5.1.3 Findings for RQ3: Perception of Safety Practice

Hypothesis hSafIsRel is rejected We perceive the results for hSafIsRel as positive because the issue of "confusing safety with reliability" raised in [22, p. 7, Assumption 1] can not be confirmed from the analysis of our responses. In fact, we observe an opposite tendency from our sample and assume this to be the effect of those SPs having been trained on that issue.

Finding 11 *The assurance of the reliability of a system does not imply the assurance of its safety.*

From the responses to qUndesiredEv-a, we derive that assured reliability of a system does not reduce the need for safety activities. Consequently, these responses do not justify the hypothesis *high reliability implies high safety*. However, we might expect to see agreement on the hypothesis *high safety implies high reliability* (qUndesiredEv-e). Likewise, our responses are ambiguous in that case. The most reasonable explanation for this ambiguity is that we did not explain the meaning of such implications along with the given answer options. Overall, the data gathered from qUndesiredEv-a–e indicates that respondents understand safety to be different from reliability. Moreover, these answers suggest that safety and safety activities are more driven by system accidents than by system failures. And this again complies with Leveson's [22] findings about the difference of safety and reliability.

Questions qEffRoleOnJob and qEffNotionOnJob 56 respondents state that their role is clearly defined. 37 perceive positive impacts on their activities, particularly, fostering clear responsibilities, authority, and escalation routes.

30 respondents state that their role is not clearly defined. 15 of them perceive negative impacts in form of unclear responsibilities, limited authority, autonomy, and space for discretionary activity as well as unclear or late decision making (Section 4.3.11).

Finding 12 *Many respondents state that their role is often not clearly defined and that they experience negative impacts from this.*

24 participants experience a non-supportive, misunderstood, or underrated safety culture. As opposed to that, 20 respondents perceive a supportive or highly-valued safety culture. 9 persons provided an ambivalent picture of safety culture, stating that they have gathered contrasting experiences (Section 4.3.12).

Finding 13 *The respondents perceive to a similar extent both, supportive and non-supportive safety cultures.*

Hypothesis hPosSelfImg is supported While responses to qValOfContrib support hPosSelfImg, the frequent indication of "medium value", particularly for qCoWorkers, suggests that some SPs might either not be convinced of the role, their profession, or even unsatisfied with their tasks and their job profile. Section 4.3.11 provides some explanation for such a dissatisfaction coming from an *unclear* role definition. Section 4.3.12 delivers an explanation from a *non-supportive* safety culture. Again, this indication provides further motivation for an extended study.

The perception of an SP's role and contribution by non-safety co-workers slightly differs from how SPs perceive their own role. This might not be too surprising because qValOfContrib and qCoWorkers redundantly measure fragments of a participant's self-perception.

Finding 14 *The respondents are self-confident about their contribution as a safety practitioner.*

5.2 Relation to Existing Evidence

In Table 8, we summarise our findings and, below, we compare them with findings from related studies.

Graaf et al. [24] identify *legacy incompatibility, lack of maturity, and additional complexity* of new methods, languages, and tools as three obstacles to the early adoption of such means. Similar obstacles were observed in software testing by Kasurinen et al. [25]. These observations are *consistent with* F1 that knowledge from previous projects has the strongest influence.

The participants' belief that FMs can have a positive impact on safety activities F3 *is strengthened* by the survey of embedded SE researchers and practitioners in [67, pp. 102f]. The authors observe that about 30% of responses from industry declare the need for FMs as a reason to adopt model-based engineering (MBE) and that MBE adoption has a positive effect on FM adoption.

Martins and Gorschek [16] observe a lack of evidence for the usefulness and usability of new approaches from safety research. Their observation is *not in conflict with* finding F3, because SPs can perceive usefulness of new FMs independent of evidence. The authors perceive a dominance of conventional approaches in practice which is *consistent with finding* F1. Furthermore, they observe a lack of investigations about the improvement of communication in the life cycle. Huber et al. [33] address this demand and confirm the findings F7 and F8. F8 indicates that such studies would be of interest to practitioners.

Chen et al. [27] observe that assurance cases can improve cross-disciplinary collaboration but are missing tool support and experienced personnel. A lack of research transfer and training could explain the *contrast to* finding F10, given that assurance cases are seen as a new method by SPs.

By observing insufficient tool support for change impact analysis in safety practice, Borg et al. [31] and De la Vara et al. [32] *contrast* the rejection of hInsufMeth.

McDermid and Rae [14] could find no satisfactory explanation to their observation that systems got "so safe despite inadequate and inadequately applied techniques." However, their assumption is *orthogonal to* finding F2, *contrasts* the rejection of hInsufMeth, and motivates further research. The lack of consensus on how to combine case-based [26] and compliance-based [28] assurance underpins this lack of clarity about the adequacy of means.

Finding F1 supports the observation of Nair et al. [30] that expert judgements and checklists are among the most frequently used references to assess safety arguments and evidence (see Figure 12). F1 is also *shared by* Rae and Alexander [44] who conclude that critical aspects of safety analysis (e.g. identifying hazards, estimating risk probability and consequence severity) often rely on expert opinion. Moreover, F1 *underpins* two out of Wang's and Wagner's [29] top ten identified decision making pitfalls.

Leveson [22] observes that safety is pervasively confused (or assumed to correlate) with reliability. The data for finding F11 *supports* her conclusion. However, the consensus of our responses suggests that there is broad awareness that safety and reliability are two related but distinct properties of a system.

Overall, we found supportive and contrasting evidence regarding most findings for RQ1, RQ2, and RQ3.

5.3 General Feedback on the Survey

The last page of our questionnaire contains a text field for general feedback. One issue, the participants criticised, pertains to *the scope and the terminology* used in the questionnaire:

The respondents noted that the inquiry is general and does not account for the diversity of safety practices in various industries. Some questions rely on a particular interpretation of safety practice leaving assumptions implicit and risking to get in conflict with other views of system safety, for example, "safety by introduction of controls" versus "safety assurance and assessment." Moreover, some of the questions are hard to answer because of a lack of standardised terminology across domains and because of missing topics, for example, legal safety requirements and regulations, human operators, socio-technical systems were not mentioned.

We see such feedback as a confirmation for the need of cross-disciplinary explorations like ours. Such explorations can help to find a terminology suitable for most SPs. When designing our questionnaire, we were driven by the variety of experiences from several collaborations with industry. Moreover, we had to prioritise and cut the question catalogue to stay within a maximum duration of 30 minutes, an amount of time we believe to be affordable by most participants. After several iterations and an email-based focus group, we finalised the questionnaire to be released.

Except for qAdeqOfMthStd and qPosImpOfFMs, the acceptably low number ($< 10\%$) of *dnk*-responses indicates that most respondents did not seem to struggle with answering. However, frequent *nand*-responses indicate difficulties in deciding on the given answer options (see, e.g. qAppOfMeth).

Another issue raised by our respondents deals with the *survey method and design* we applied:

Some questions may motivate one to answer in a particular way and solicit specific support. LIKERT scales impose an abstraction with the risk to deny more accurate answers such as "I often highly agree and sometimes I strongly disagree." Moreover, LIKERT scales should be substituted by open questions more

Table 8: Overview of findings from hypothesis analysis

RQ1: Which means do SPs typically rely on? How helpful are those means to them?	**RQ2:** Which typical process factors have influence on SPs' decisions & performance?	**RQ3:** How do SPs perceive and understand their role in the process or life cycle?
F1: The responses suggest that safety strongly depends on expert opinion and project memory. F2: In the considered high automation domains, the adequacy of standards and methods for safety assurance is rated with low to very low. F3: Among the informed respondents, formal methods are believed to be beneficial.	F4: The respondents suggest that resources occasionally but not typically govern SPs' performance. F5: Our respondents suggest that safety is not typically a question of cost-benefit. F6: A lack of collaboration downgrades the performance of safety activities. F7: Interaction between safety and security practitioners rarely occurs in requirements and assurance activities. F8: Early collaboration of safety and security experts is viewed as essential. F9: Our respondents see diverse expertise as essential for safety practitioners. F10: Respondents from the automotive domain perceive adaptation of skills to new technologies to be more relevant for SPs than respondents from the aeronautics domain.	F11: The assurance of the reliability of a system does not imply the assurance of its safety. F12: Many respondents state that their role is often not clearly defined and that they experience negative impacts from this. F13: The respondents perceive to a similar extent both, supportive and non-supportive safety cultures. F14: The respondents are self-confident about their contribution as a safety practitioner.

appropriate for exploratory studies where the construct is not known or (entirely) fixed beforehand.

To compensate for this issue, we allowed open answers and present results from their qualitative analysis (e.g. in the Sections 4.3.1, 4.3.2, 4.3.9, 4.3.11 and 4.3.12). More open questions reduce the risks of bias and constrained data acquisition. However, it is worth noting that, as opposed to interviews, too many open answers in large-scale questionnaires can also be demanding for the respondents and, thus, lead to a high number of partial data points.

5.4 Validity Procedure after Survey Execution

Here, we assess our survey design with respect to internal and external validity as well as reliability [62, 68].

5.4.1 Internal Validity

To reduce threats to internal validity, we followed recommendations on questionnaire-based surveys in the software and systems engineering domain [69]. Section 5.3 discusses arguments for internal validity as a response to the general feedback on our survey. Additionally, the everyday use of English among the majority of survey participants (see the criterion Native Languages) supports the accuracy of a large fraction of the data points.

5.4.2 External Validity

To which extent would the procedure in Section 3 lead to similar results with different samples?

Our sampling procedure is network-guided and, hence, not uniformly random [60]. Section 4.2 shows that our sample varies over the scales of all classification criteria (Table 1). The overlap of the summer holiday season with the sampling period might have slightly influenced the total number of data points. Regarding *safety cultures*, our sample well represents the backgrounds (mainly engineering), domains (mainly aerospace, automotive, railway), and geographical regions (mainly Europe and US) identified in Section 4.2. Moreover, the rejection of hInsufMeth:EngDifSci (i.e., practitioners differ from academics in their view of inadequacy of means) strongly reduces the likelihood that the participation of researchers biases the results towards an academic view.

According to Figure 11, 19 out of 124 respondents stated that they have been working on safety-related topics as a *researcher in academia*, that is, the role or responsibility profile which we associate the least of all with genuine practical experience. *Only 4 of them declared to be solely academic researchers.* 8 stated to be SPs, too; 7 have also done research in industry; 11 have worked as software, systems, requirements, reliability, or health & safety practitioners in addition. This again strengthens our belief that our results are not biased towards an academic view.

In comparison with focus groups and individual interviews, anonymous on-line surveys are highly valuable inasmuch as they help mitigating the two following risks:

1. In collaborations between academia and industry, industrial participants are likely from the management, or senior engineers, or research engineers not necessarily regularly connected to the operational teams. Such collaborations bear the risk of the sample getting biased towards these roles. With an on-line survey advertised on multiple channels, we are convinced to have reduced such bias.
2. Safety activities are legally critical. The authors' experience from personal interviews is that practitioners tend to avoid talking loosely about their organisations and, where aggravating, to moderately generalise. Our impression from the respondents' occasionally quite open comments makes us to believe that the risk of this bias is lower in anonymous surveys. Note that subjectivity has to be handled by other means in both questionnaires and interviews.

Leveson [22, p. 211] states that FMEA, with its limited applicability for safety analysis, is less frequently used as a hazard analysis technique than FTA or ETA. As opposed to Leveson's observation, our respondents most often state that they worked with FMEA-based techniques (cf. Figure 8). One reason for this discrepancy could be that we provided a small set of techniques as answer options to Familiarity with Methods (e.g. ETA was not included). Assuming that respondents are reluctant to use the "Other" field, this might have led to a bias towards the specified options. Assuming that Leveson's observation is mainly derived from US safety cultures, this discrepancy could also result from the focus of our sample on European safety cultures (cf. Geographical Regions in Section 4.2). While this issue limits the external validity, we believe that the results for the questions and the hypotheses are not harmed.

The isolation of most questions allows a per-question analysis. Particularly, the 59 partial responses do not affect any complete data points and thus were taken into consideration for the questions for which they delivered responses (cf. variation of N values). The relatively high number of registered views (565) might stem from users checking the questionnaire start page and concluding that they do not belong to the target group (Section 3.5): Diverse preconceptions of safety, diverse channel members, and necessarily short advertisements might have played a role. While this issue reduces the total number of relevant respondents, we believe to have well prevented illegible respondents from participating.

We expect the population of SPs to be 2 to 3 orders of magnitude larger than our sample ($N = 91, N' = 124$). According to an estimation in [70], general conclusions require about twice as many data points. For example, Manotas et al. [64] sample the population with the support of global software companies. Such approaches might lead to a larger number of responses than with using volunteer and cluster-based sampling from on-line discussion channels. However, we believe our approach better reduces biases to a specific domain, company, or region.

5.4.3 *Reliability*

To which extent would a repetition of the procedure in Section 3 with the same sample lead to the same results?

It is difficult to exactly repeat this survey in the short term because our advertisements covered many of the relevant on-line channels and we expect many of the respondents not willing to participate again within short-term. This is a general problem for studies of this kind. Therefore, we suggest to 1. provide incentives, 2. pursue off-line channels as well, 3. repeat the study in the long term, and 4. extend the sampling period. For example, Mendez-Fernandez et al. [71] provide a longitudinal design supporting repeatability and hence the determination of reliability of the results.

5.5 *Lessons Learned*

Regarding the sample size (Section 4.2), we wished to get more responses against the background of the effort we spent in reaching out to the population (Section 4.1). From the Unipark questionnaire view statistics, we saw that in some of the larger discussion forums, users seemed to appear noticeably reluctant to respond. The return rates estimated in Section 4.1 are low. In some channels, our friendly, singular, and topic-related post was even penalised by deleting the post or by loosing channel membership. Unfortunately, non-commercial panels, such as SoSciSurvey[12] and SurveyCircle (Table 6), do not offer profiling facilities to focus on engineering professionals. In the case of no budget for incentives and for paying commercial panels, these circumstances make it very difficult for empirical researchers to get a representative sample.

6 Conclusions and Future Work

We conducted a questionnaire-based cross-sectional on-line survey of safety practitioners. Our objective was to explore safety practice by asking practitioners about means they rely on, process factors influencing their work, and their role in the life cycle, and by checking several observations from previous collaborations.

6.1 *Summary of Findings and Implications*

Below, observations marked with + represent our aspirations when performing the study. Observations marked with – represent our apprehensions. Other items accommodate neutral observations.

We collected evidence in support of several hypotheses leading to the following observations:

- Our respondents confirm that safety decision making is largely based on expert opinion and experience from previous projects.
- For connected systems (e.g. systems of systems, connected transport systems), our respondents believe that assurance of safety also relies on the assurance of IT security.
- \+ They see a clear benefit in the interaction of safety and security activities. We support the agenda in [16] and encourage further research on integrated safety/security approaches.
- \+ The survey participants believe that formal methods have a positive impact on safety activities.
- – Many of them see currently applied standards and methods as inadequate for the assurance of technologies (e.g. adaptive control, machine learning) in high automation and autonomy. This justifies the question of whether future systems will be safe enough and why this would be the case [14]?

Our respondents allow further observations:

- They suggest that resources occasionally govern safety practitioners' decisions and performance but also that safety is not typically handled as a cost/benefit question.
- The survey participants refrain from seeing safety as a consequence of reliability. This stands in contrast with Leveson's former observation that safety is pervasively confused with reliability [22].
- – The respondents believe that many of their non-safety co-workers' share at most medium appreciation of safety practitioners' contributions to the life cycle.

[12] See https://www.soscisurvey.de.

- The respondents are indecisive on whether the conventional or ready-to-use methods they (could) apply scale sufficiently.

The last finding motivates analysis along the lines of McDermid and Rae [14]. If we are unsure about whether means have become inadequate and, as found for safety RE in [16], if conventional approaches are dominant and we lack evidence for efficacy of novel research, what good is safety research for safety practitioners?

In summary, we share with Alexander and Rae [49] the impression that *empirical research in system safety is still in an early stage*, on the one hand, offering many opportunities to perform cross-disciplinary studies and, on the other hand, bearing large risks of not exactly knowing to which extent safety practitioners are applying state of the art and able to do their best. We believe that this is a serious issue to be tackled by software and system safety research.

6.2 Future Work

We seek to extend our analysis by revisiting findings from the collected data set and not discussed in this work. Furthermore, we are going to identify and evaluate further hypotheses and ask more *why*- and *how*-questions.

Aspiring to the exploratory approach and grounded theory [62, p. 298], we can further engage with our survey participants using the *focus group method* [72], request for comments on our findings, and ask them for approaches to overcome the identified issues. Additionally, we refine our construct and elaborate on a subset of questions to investigate the applicability of formal methods like, for example, shown in [66].

Our research design can be extended towards the application of the *goal question metric* approach [73]: The results of the hypotheses analysis promotes the definition of *goals* of safety activities, the *survey questions* corresponding to the hypotheses can be refined, and process and product *metrics* be derived from the refined questions, for example, as already suggested and discussed in [74, 75]. Our study object includes safety practitioners and, consequently, some of these metrics get measurable by questionnaires.

Our results are a good starting point for the design of a longitudinal study, offering possibilities to identify and validate causal relationships among the measured variables (Table 2). Inspired by [30] and [76], we want to adapt our research design to support investigations of *confirmation bias* in assurance arguments [49, 77].

Declaration of Interest

None.

Acknowledgments

The first author of this work is partly supported by the Deutsche Forschungsgemeinschaft (DFG) under the Grant no. 381212925. It is our pleasure to thank all survey participants for their valuable responses, and several practitioners, researchers, and students for acting as pilot run respondents and for providing us with initial feedback. Special thanks go to Mohammed Hussein and Dai Yang whose analyses yielded important preliminary findings for initiating this survey. We are indebted to Martin Wildmoser for attending the final interview for [4] with friendly support of the Validas AG[13] in Munich and to further enthusiastic safety experts from various German industries for participation in the interviews for [5]. Technical University of Munich (TUM) and University of York have been excellent working environments. I would like to thank Manfred Broy for his senior advice and for providing the research infrastructure. Daniel Mendez-Fernandez deserves cordial gratitude for giving us Unipark advice and granting us access to this platform using the TUM Informatics faculty license.

REFERENCES

[1] Charles Perrow. *Normal Accidents: Living with High-Risk Technologies*. Basic Books, NY, USA, 1984.

[2] J.N Sorensen. Safety culture: a survey of the state-of-the-art. *Reliability Engineering & System Safety*, 76(2):189 – 204, 2002.

[3] Rafiq M. Choudhry, Dongping Fang, and Sherif Mohamed. The nature of safety culture: A survey of the state-of-the-art. *Safety Science*, 45(10):993 – 1012, 2007.

[4] M. Hussein. Current challenges of system safety practitioners: Qualitative analysis of on-line discussions. Unpublished master's thesis, Technical University of Munich, 2016.

[5] Dai Yang. Hazards from high system entropy: An explorative analysis of case reports. Unpublished master's thesis, Technical University of Munich, 2016.

[6] Albert Endres. An analysis of errors and their causes in system programs. *IEEE Transactions on Software Engineering*, SE-1(2):140–149, 6 1975.

[7] Robyn R. Lutz. Analyzing software requirements errors in safety-critical, embedded systems. In *IEEE Int. Symp. Req. Eng.*, pages 126–33. IEEE, 1993.

[8] Health and Safety Executive. *Out of Control*. HSE Books, 2003.

[9] Marko Wolf and Robert Lambert. Hacking trucks - cybersecurity risks and effective cybersecurity protection for heavy duty vehicles. In Peter Dencker, Herbert Klenk, Hubert B. Keller, and Erhard Pldererder, editors, *Automotive - Safety & Security 2017 - Sicherheit und Zuverlssigkeit fr automobile Informationstechnik*, pages 45–60. Gesellschaft fr Informatik, Bonn, 2017.

[10] Alex Wright. Hacking cars. *Communications of the ACM*, 54(11):18, nov 2011.

[11] Frederick P. Brooks, Jr. *The Mythical Man Month: Essays on Software Engineering*. Addison-Wesley Longman, Amsterdam, 20th anniversary edition, Aug 1995.

[12] John C. Knight. Safety critical systems: Challenges and directions. In *Proceedings of the 24th International Conference on Software Engineering*, ICSE '02, pages 547–50, New York, NY, USA, 2002. ACM.

[13] Peter G. Neumann. Risks to the public. *ACM SIGSOFT Software Engineering Notes*, 43(2):8–11, may 2018.

[14] J. A. McDermid and A. J. Rae. How did systems get so safe without adequate analysis methods? In *9th IET International Conference on System Safety and Cyber Security (2014)*, pages 1–6. Institution of Engineering and Technology, 2014.

[15] Leon J. Osterweil. Be gracious. *ACM SIGSOFT Software Engineering Notes*, 43(2):4–6, may 2018.

[13] See http://www.validas.de.

[16] Luiz Eduardo G. Martins and Tony Gorschek. Requirements engineering for safety-critical systems: A systematic literature review. *Information and Software Technology*, 75:71–89, jul 2016.

[17] R. D. Alexander, A. J. Rae, and M. Nicholson. Matching research goals and methods in system safety engineering. In *5th IET International Conference on System Safety 2010*, pages 1–8, Oct 2010.

[18] A. J. Rae, M. Nicholson, and Robert D. Alexander. The state of practice in system safety research evaluation. In *5th IET International Conference on System Safety 2010*, pages 1–8. IET, 2010.

[19] Ricardo Valerdi and Heidi L. Davidz. Empirical research in systems engineering: challenges and opportunities of a new frontier. *Systems Engineering*, 12(2):169–181, 2009.

[20] J. C. Laprie. *Dependability: Basic Concepts and Terminology*. Springer, Berlin/New-York, 1992.

[21] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr. Basic concepts and taxonomy of dependable and secure computing. *Dependable and Secure Computing, IEEE Transactions on*, 1(1):11–33, 1 2004.

[22] Nancy Gail Leveson. *Engineering a Safer World: Systems Thinking Applied to Safety*. Engineering Systems. MIT Press, 1 2012.

[23] Tom Dwyer. Industrial safety engineering–challenges of the future. *Accident Analysis & Prevention*, 24(3):265–273, 1992.

[24] B. Graaf, M. Lormans, and H. Toetenel. Embedded software engineering: the state of the practice. *IEEE Software*, 20(6):61–69, Nov 2003.

[25] Jussi Kasurinen, Ossi Taipale, and Kari Smolander. Software test automation in practice: Empirical observations, 2010.

[26] John Hatcliff, Alan Wassyng, Tim Kelly, Cyrille Comar, and Paul Jones. Certifiably safe software-dependent systems: challenges and directions. In *Proceedings of the on Future of Software Engineering - FOSE 2014*, pages 182–200. ACM Press, 2014.

[27] Jinghui Chen, Micayla Goodrum, Ronald Metoyer, and Jane Cleland-Huang. How do practitioners perceive assurance cases in safety-critical software systems? In *Proceedings of the 11th International Workshop on Cooperative and Human Aspects of Software Engineering - CHASE'18*, pages 57–60. ACM Press, 2018.

[28] Andrea Ceccarelli and Nuno Silva. Analysis of companies gaps in the application of standards for safety-critical software. In Floor Koornneef and Coen van Gulijk, editors, *Computer Safety, Reliability, and Security*, pages 303–313, Cham, 2015. Springer International Publishing.

[29] Yang Wang and Stefan Wagner. On groupthink in safety analysis. In *Proceedings of the 40th International Conference on Software Engineering Software Engineering in Practice - ICSE-SEIP'18*, pages 266–275. ACM Press, 2018.

[30] Sunil Nair, Jose Luis de la Vara, Mehrdad Sabetzadeh, and Davide Falessi. Evidence management for compliance of critical systems with safety standards: A survey on the state of practice. *Information and Software Technology*, 60:1 – 15, 2015.

[31] Markus Borg, Jose Luis de la Vara, and Krzysztof Wnuk. Practitioners' perspectives on change impact analysis for safety-critical software – a preliminary analysis. In *Lecture Notes in Computer Science*, pages 346–358. Springer International Publishing, 2016.

[32] Jose Luis de la Vara, Markus Borg, Krzysztof Wnuk, and Leon Moonen. An industrial survey of safety evidence change impact analysis practice. *IEEE Transactions on Software Engineering*, 42(12):1095–1117, dec 2016.

[33] Michael Huber, Michael Brunner, Clemens Sauerwein, Carmen Carlan, and Ruth Breu. Roadblocks on the highway to secure cars: An exploratory survey on the current safety and security practice of the automotive industry. In *Developments in Language Theory*, pages 157–171. Springer International Publishing, 2018.

[34] Arlene G. Fink. *How to Conduct Surveys: A Step-By-Step Guide*. SAGE, 2016.

[35] Andreas Jedlitschka, Marcus Ciolkowski, and Dietmar Pfahl. Reporting experiments in software engineering. In *Guide to Advanced Empirical Software Engineering*, pages 201–228. Springer London, 2008.

[36] Barbara A. Kitchenham and Shari L. Pfleeger. *Personal Opinion Surveys*, chapter 3, pages 63–92. Springer, 2008.

[37] Barbara Kitchenham, Hiyam Al-Khilidar, Muhammed Ali Babar, Mike Berry, Karl Cox, Jacky Keung, Felicia Kurniawati, Mark Staples, He Zhang, and Liming Zhu. Evaluating guidelines for reporting empirical software engineering studies. *Empirical Software Engineering*, 13(1):97–121, oct 2007.

[38] Mario Gleirscher. *Behavioral Safety of Technical Systems*. Dissertation, Technische Universität München, 2014. .

[39] Anne Ruguru Nyokabi. Practical safety challenges: An online survey of safety practitioners demands, problems and expectations. Unpublished master's thesis, Technical University of Munich, 2017.

[40] Anne Nyokabi and Mario Gleirscher. System safety practice: An interrogation of practitioners about their activities, challenges, and views with a focus on the European region – questionnaire, 2017. To be archived with Xenodo.

[41] Juliet M. Corbin and Anselm L. Strauss. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage, 4 edition, 11 2015.

[42] Robin Bloomfield and Peter Bishop. Safety and assurance cases: Past, present and possible future – an adelard perspective. In *Making Systems Safer*, pages 51–67. Springer London, dec 2009.

[43] T. C. Lethbridge, J. Singer, and A. Forward. How software engineers use documentation: the state of the practice. *IEEE Software*, 20(6):35–39, Nov 2003.

[44] Andrew Rae and Rob Alexander. Forecasts or fortune-telling: When are expert judgements of safety risk valid? *Safety Science*, 99:156–165, nov 2017.

[45] Elizabeth Bjarnason, Per Runeson, Markus Borg, Michael Unterkalmsteiner, Emelie Engstrm, Bjrn Regnell, Giedre Sabaliauskaite, Annabella Loconsole, Tony Gorschek, and Robert Feldt. Challenges and practices in aligning requirements with verification and validation: a case study of six companies. *Empirical Software Engineering*, 19(6):1809–1855, jul 2013.

[46] Ian Dodd and Ibrahim Habli. Safety certification of airborne software: An empirical study. *Reliability Engineering & System Safety*, 98(1):7–23, feb 2012.

[47] Erin Lim, Nitin Taksande, and Carolyn Seaman. A balancing act: What software practitioners have to say about technical debt. *IEEE Software*, 29(6):22–27, nov 2012.

[48] Christoph K. Streb. *Exploratory Case Study*, pages 370–371. SAGE, 2010.

[49] Andrew John Rae and Rob D. Alexander. Probative blindness and false assurance about safety. *Safety Science*, 92:190–204, 2017.

[50] T. Cant. System safety: Where next? In *System Safety Conference incorporating the Cyber Security Conference 2013, 8th IET International*, pages 1–10, 10 2013.

[51] Alessia Knauss, Jan Schröder, Christian Berger, and Henrik Eriksson. Paving the roadway for safety of automated vehicles: An empirical study on testing challenges. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 1873–80. IEEE, jun 2017.

[52] John Knight. *Fundamentals of Dependable Computing for Software Engineers*. Chapman & Hall/CRC Innovations in Software Engineering and Software Development. Chapman and Hall/CRC, 2012.

[53] R. E. Bloomfield, P. K. D. Froome, and B. Q. Monahan. Formal methods in the production and assessment of safety critical software. *Reliability Eng. & Sys. Safety*, 32(1-2):51–66, 1991.

[54] Leonor M. Barroca and John A. McDermid. Formal methods: Use and relevance for the development of safety-critical systems. *Comp. J.*, 35(6):579–99, 1992.

[55] Jim Woodcock, Peter Gorm Larsen, Juan Bicarregui, and John Fitzgerald. Formal methods: Practice and experience. *ACM Comput. Surv.*, 41(4):19:1–19:36, 10 2009.

[56] J. Lockhart, C. Purdy, and P. Wilsey. Formal methods for safety critical system specification. In *IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 201–204, Aug 2014.

[57] Stephen Checkoway, Damon McCoy, Brian Kantor, Danny Anderson, Hovav Shacham, Stefan Savage, Karl Koscher, Alexei Czeskis, Franziska Roesner, and Tadayoshi Kohno. Comprehensive experimental analyses of automotive attack surfaces. In *20th USENIX Security Symposium, San Francisco, CA, USA, August 8-12, 2011, Proceedings*, pages 1–16, 2011.

[58] Melvin E. Conway. How do commitees invent? *Datamation*, 14(4):28–31, April 1968. Design Organization Criteria.

[59] M. Napolano, F. Machida, R. Pietrantuono, and D. Cotroneo. Preventing recurrence of industrial control system accident using assurance case. In *Software Reliability Engineering Workshops (ISSREW), 2015 IEEE International Symposium on*, pages 182–189, Nov 2015.

[60] Haslam and McGarty. *Research Methods and Statistics in Psychology*. Routledge, 5 edition, 2009.

[61] Kimberly A. Neuendorf. *The Content Analysis Guidebook*. Sage, 2 edition, 8 2016.

[62] Forrest Shull, Janice Singer, and Dag I. K. Sjøberg, editors. *Guide to Advanced Empirical Software Engineering*. Springer, Oct 2008.

[63] Younghwa Lee, Kenneth A. Kozar, and Kai R. T. Larsen. The technology acceptance model: Past, present, and future. *Comm. AIS*, 12:752–80, 2003.

[64] I. Manotas, C. Bird, R. Zhang, D. Shepherd, C. Jaspan, C. Sadowski, L. Pollock, and J. Clause. An empirical study of practitioners' perspectives on green software engineering. In *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*, pages 237–248, May 2016.

[65] Naomi B. Robbins and Richard M. Heiberger. Plotting likert and other rating scales. In *Joint Statistical Meeting*, pages 1058–66, 2011.

[66] Mario Gleirscher and Diego Marmsoler. Formal methods in dependable systems engineering: A survey of professionals from Europe and North America. Unpublished working paper, Department of Computer Science, University of York, 2018.

[67] Grischa Liebel, Nadja Marko, Matthias Tichy, Andrea Leitner, and Jrgen Hansson. Model-based engineering in the embedded systems domain: an industrial survey on the state-of-practice. *Software & Systems Modeling*, 17(1):91–113, mar 2016.

[68] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslén. *Experimentation in Software Engineering*. Springer, 6 2012.

[69] Marcus Ciolkowski, Oliver Laitenberger, Sira Vegas, and Stefan Biffl. Practical experiences in the design and conduct of surveys in empirical software engineering. In Reidar Conradi and Alf Inge Wang, editors, *Empirical Methods and Studies in Software Engineering: Experiences from ESERNET*, pages 104–128, Berlin, Heidelberg, 2003. Springer.

[70] Mario Gleirscher, Simon Foster, and Jim Woodcock. New opportunities for integrated formal methods. Unpublished working paper, Department of Computer Science, University of York, 2019. 75%.

[71] Daniel Méndez-Fernández, Stefan Wagner, Marcos Kalinowski, Michael Felderer, Priscilla Mafra, Antonio Vetro, Tayana Conte, M.-T. Christiansson, Des Greer, Casper Lassenius, Tomi Männistö, M. Nayabi, Markku Oivo, Birgit Penzenstadler, Dietmar Pfahl, Rafael Prikladnicki, Günther Ruhe, André Schekelmann, S. Sen, Rodrigo O. Spínola, Ahmet Tuzcu, Jose Luis de la Vara, and Roel Wieringa. Naming the pain in requirements engineering - contemporary problems, causes, and effects in practice. *Empirical Software Engineering*, 22(5):2298–2338, 2017.

[72] J. Kontio, L. Lehtola, and J. Bragge. Using the focus group method in software engineering: obtaining practitioner and user experiences. In *Empirical Software Engineering (ISESE). Int. Symposium on*, pages 271–280, Aug 2004.

[73] Victor Basili, Gianluigi Caldiera, and Dieter H. Rombach. The goal question metric approach. In J. Marciniak, editor, *Encyclopedia of Software Engineering*. Wiley, 1994.

[74] John Murdoch, Graham Clark, Antony Powell, and Paul Caseley. Measuring safety: Applying psm to the system safety domain. In *Proceedings of the 8th Australian Workshop on Safety Critical Systems and Software*, volume 33 of *SCS '03*, pages 47–55, Darlinghurst, Australia, Australia, 2003. Australian Computer Society, Inc.

[75] Yaping Luo and Mark van den Brand. Metrics design for safety assessment. *Information and Software Technology*, 73:151–163, may 2016.

[76] Mario Gleirscher and Carmen Carlan. Arguing from hazard analysis in safety cases: A modular argument pattern. In *High Assurance Systems Engineering (HASE), 18th Int. Symp.*, pages 53–60, 1 2017.

[77] Nancy Leveson. The use of safety cases in certification and regulation. *Journal of System Safety*, 47(6):e–Edition, 2011.

A Summary of All Responses

The following tables present data summaries for all closed (q)uestions according to Table 4 and questions for classification according to Table 1. The "Option" column refers to the parts (if any) of multi-part questions. The "NA's" column signifies the number of invalid data points for each (part of a) question. The checksum (including invalid responses) of each row results in $N = 152$ responses. Rows with NA's = 0 result from parts (i.e., answer categories) added after content analysis of half-open questions (Section 3.6.1). The questions qEffRoleOnJob and qEffNotionOnJob are open and, hence, not accompanied by a corresponding table.

qValueOfKnow	Value					
Option / N	Very low	Low	Medium	High	Very high	NA's
a / 96	1	4	14	40	37	56
b / 96	1	3	25	36	31	56
c / 95	2	10	33	37	13	57
d / 95	1	1	13	53	27	57
e / 96	9	34	37	10	6	56
f / 96	1	8	47	32	8	56
g / 97	1	0	7	39	50	55

Legend: a. Hazard list from previous projects, b. Case (accident, incident) reports, c. Inspection checklist, d. Expert opinions, e. Management recommendations, f. Co-workers' recommendations, g. Safety-related project experience

qImpOfConstr	Impact					
Option / N	Do not know	No impact	Low impact	Medium impact	High impact	NA's
a / 98	5	5	12	39	37	54
b / 96	2	4	6	31	53	56
c / 95	6	2	5	36	46	57
d / 95	5	4	12	31	43	57
e / 97	4	3	9	25	56	55
f / 97	1	4	23	35	34	55
g / 98	2	2	11	27	56	54

Legend: a. Budget cuts, b. Late or unclear choice of safety concepts, c. Postponed safety decisions, d. Schedule pressure, e. Erroneous hazard analyses, f. Vague safety standards, g. Inexperienced safety engineers

qImpOfEco	Frequency				
Option / N	Often	Rarely / Occasionally	Never	Do not know	NA's
– / 99	36	48	9	6	53

qAdeqOfMthStd	Adequacy					
Option / N	Do not know	Not adequate	Slightly adequate	Adequate	Very adequate	NA's
a / 100	32	31	26	11	0	52
b / 101	22	30	33	13	3	51
c / 101	26	31	27	16	1	51
d / 101	22	48	23	8	0	51
e / 100	36	7	25	27	5	52
f / 101	29	10	26	30	6	51
g / 99	32	28	21	17	1	53

Legend: a. Self-adaptive systems, b. Highly automated and autonomous driving, c. Distributed networked systems, d. AI/ML-based applications, e. Medical and healthcare applications, f. Highly automated air traffic control, g. Consumer or commercial drones

qAppOfMeth	Agreement						
Option / N	Do not know	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	NA's
– / 102	5	3	28	18	37	11	50

qPosImpOfFMs	Impact					
Option / N	Do not know	No impact	Low impact	Medium impact	High impact	NA's
– / 62	4	3	15	23	17	90

qImprOfSkills	Agreement						
Option / N	Do not know	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	NA's
a / 96	1	1	3	5	37	49	56
b / 95	9	1	14	17	36	18	57
c / 96	6	1	4	13	53	19	56
d / 95	12	6	18	21	32	6	57

Legend: a. Adapt skills to new technologies, b. Study state-of-the-art safety principles, c. Juniors learn from seniors, d. Juniors learn from accident reports

qIntOfSafSec	Agreement						
Option / N	Do not know	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	NA's
a / 94	3	5	13	14	35	24	58
b / 95	3	10	30	23	19	10	57
c / 93	4	4	11	15	36	23	59
d / 93	6	8	22	28	20	9	59
e / 93	5	3	10	9	42	24	59
f / 94	11	1	9	13	43	17	58
g / 92	11	0	12	16	41	12	60
h / 94	4	1	4	12	38	35	58
i / 94	5	4	3	12	49	21	58
j / 94	3	0	3	5	41	42	58

Legend: a. Security is prerequisite for safety, b. Safety is prerequisite for security, c. SPs depend on security practitioners, d. Security practitioners depend on SPs, e. Safety assurance requires security assurcance, f. Rare interaction in requirements stage, g. Rare interaction in assurance stage, h. Lack of collaboration is hazardous, i. Lack of collaboration is inefficient, j. Involvement in RE improves safety

qNotionOfSafety	Multiple Choice		
Option / N	Checked	Unchecked	NA's
a / 95	30	65	57
b / 95	27	68	57
c / 95	50	45	57
d / 95	30	65	57
e / 95	21	74	57

Legend: a. A cost factor, b. A beneficial factor, c. A necessity independent of cost, d. A tedious mandated task, e. A secondary issue

qPrioOfSafety	Agreement						
Option / N	Do not know	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	NA's
a / 97	1	1	1	7	44	43	55
b / 97	3	2	3	8	40	41	55
c / 97	2	2	2	16	44	31	55
d / 97	2	2	1	7	41	44	55

Legend: a. Safety is given high priority, b. Management highly values safety, c. SPs have declared authority, d. Safety process is defined

qUndesiredEv	Agreement						
Option / N	Do not know	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	NA's
a / 97	2	57	28	3	5	2	55
b / 96	1	11	18	12	37	17	56
c / 96	1	4	6	15	44	26	56
d / 97	2	17	18	21	29	10	55
e / 96	3	25	25	15	19	9	56

Legend: a. Lack of failures reduces need for safety, b. Accidents drive need for safety, c. Accidents help SPs argue for safety, d. Lack of accidents reduces need for safety, e. Safety implies reliability

qValOfContrib	Value					
Option / N	Very low	Low	Medium	High	Very high	NA's
– / 95	2	6	25	48	14	57

qCoWorkers	Value					
Option / N	Very low	Low	Medium	High	Very high	NA's
– / 95	4	19	37	28	7	57

qImpOfExp	Agreement						
Option / N	Do not know	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	NA's
a / 95	3	2	6	11	41	32	57
b / 95	3	1	4	13	47	27	57
c / 96	16	2	2	27	29	20	56

Legend: a. Senior SPs outperform junior SPs, b. Previous projects experience is beneficial, c. Adversarial thinking improves hazard analysis

Educational Background	Multiple Choice (Classification)		
Option / N	Checked	Unchecked	NA's
a / 124	49	75	28
b / 124	13	111	28
e / 124	23	101	28
f / 124	31	93	28
g / 124	22	102	28
h / 124	12	112	28
k / 124	6	118	28

Legend: a. Computer Science, b. Systems Engineering, e. Safety Science, f. Electrical and Electronics Engineering, g. Mechanical and Aerospace Engineering, h. Physics and Mathematics, k. Other Discipline

Application Domains	Multiple Choice (Classification)		
Option / N	Checked	Unchecked	NA's
a / 124	51	73	28
b / 124	54	70	28
c / 124	16	108	28
d / 124	19	105	28
e / 124	30	94	28
f / 124	17	107	28
g / 124	24	100	28
h / 124	9	115	28
j / 124	25	99	28
l / 124	15	109	28
m / 124	8	116	28
o / 152	15	137	0
p / 152	6	146	0

Legend: a. Automotive and Transport Systems, b. Aerospace Industry, c. IT Infrastructure and Networking, d. Power and Nuclear Industry, e. Industrial Processes and Plant Automation, f. Electronic Devices and Appliances, g. Healthcare Systems, h. Construction and Building Automation, j. Industrial Machinery, l. Naval Systems, m. Other Domain, o. Railway and Cablecar Systems, p. Military and Defense Systems

Level of Experience	Single Choice (Years Of Experience In Levels)					
Option / N	< 3	3 - 7	8 - 15	16 - 25	> 25	NA's
– / 119	27	27	27	18	20	33

Familiarity with Standards		Multiple Choice (Classification)	
Option / N	Checked	Unchecked	NA's
a / 124	70	54	28
b / 124	19	105	28
c / 124	47	77	28
e / 124	3	121	28
f / 124	47	77	28
h / 124	14	110	28
i / 124	8	116	28
k / 152	3	149	0
l / 152	3	149	0
m / 152	12	140	0
n / 152	4	148	0
o / 152	14	138	0

Legend: a. Generic, b. Machinery, c. Automotive and Transport, e. Agriculture, f. Aerospace and Avionics, h. Not Familiar, i. Other Standard, k. Nuclear and Other Energy, l. Medical Devices, m. Railway, n. Methodology and Tooling, o. Military and Defense

Familiarity with Methods		Multiple Choice (Classification)	
Option / N	Checked	Unchecked	NA's
a / 124	90	34	28
b / 124	78	46	28
c / 124	16	108	28
d / 124	59	65	28
e / 123	4	119	29
f / 124	15	109	28
g / 124	15	109	28
i / 152	9	143	0
j / 152	3	149	0
k / 152	5	147	0
l / 152	3	149	0
m / 152	5	147	0
n / 152	4	148	0
o / 152	6	146	0
p / 152	3	149	0

Legend: a. Failure Mode Effects, b. Fault Trees, c. STAMP-based Methods, d. Hazard Operability, e. Security Threats, f. Not Familiar, g. Other Concept/Method, i. Functional and Design Risk, j. Probabilistic Risk, k. Assurance Cases, l. Root Causes, m. Event Trees, n. Automated VandV, o. Dependencies and Interactions, p. Bidirectional Methods

Native Languages		Multiple Choice (Classification)	
Option / N	Checked	Unchecked	NA's
a / 124	54	70	28
b / 124	38	86	28
c / 124	5	119	28
d / 124	6	118	28
e / 124	2	122	28
g / 152	3	149	0
h / 152	3	149	0
i / 152	14	138	0

Legend: a. English, b. German, c. Italian, d. French, e. Chinese, g. Portuguese, h. Swedish, i. Other language

Working Languages		Multiple Choice (Classification)	
Option / N	Checked	Unchecked	NA's
a / 124	113	11	28
b / 124	41	83	28
c / 124	5	119	28
d / 124	10	114	28
e / 124	12	112	28

Legend: a. English, b. German, c. Italian, d. French, e. Other language

Safety-related Roles		Multiple Choice (Classification)	
Option / N	Checked	Unchecked	NA's
a / 124	62	62	28
e / 124	29	95	28
f / 124	9	115	28
g / 124	39	85	28
h / 124	23	101	28
j / 124	11	113	28
k / 124	3	121	28
m / 66	12	54	86
n / 124	4	120	28
q / 152	3	149	0

Legend: a. Safety Practitioner, e. Software Practitioner, f. Electrical or Electronics Practitioner, g. Systems Practitioner, h. Requirements Practitioner, j. Health and Safety Practitioner, k. IT Security Practitioner, m. Reliability Practitioner, n. Other Role, q. V and V Practitioner

Safety-related Roles		Multiple Choice (Classification)	
Option / N	Checked	Unchecked	NA's
b / 124	26	98	28
c / 124	19	105	28
d / 124	5	119	28
l / 76	13	63	76
p / 152	12	140	0

Legend: b. Researcher in Industry, c. Researcher in Academia, d. Assistant, Trainee, or Junior, l. Practitioner with PhD degree, p. Consultant / Assessor

Table 9: List of abbreviations used in this article

AC	Acceptance Criterion
CAST	Causal Analysis using System Theory
CCA	Common Cause Analysis
CIA	Change Impact Analysis
CMA	Common Mode Analysis
DV	Dependent Variable
ETA	Event Tree Analysis
FFA	Functional Failure Analysis
FHA	Functional Hazard Analysis
FM	Formal Method
FMEA	Failure Mode Effects Analysis
FMECA	Failure Mode, Effects, and Criticality Analysis
FMEDA	Failure Mode, Effects, and Diagnostic Analysis
FRAM	Functional Resonance Analysis Method
FTA	Fault Tree Analysis
GSN	Goal Structuring Notation
HazOp	Hazard Operability (studies)
HRA	Health Risk Assessment
IV	Independent Variable
MC	Multiple Choice
NA	Not Available
OSHA	Operation & Support Hazard Analysis
PHA	Preliminary Hazard Analysis
PHL	Preliminary Hazard List
RCA	Root Cause Analysis
RE	Requirements Engineering
RQ	Research Question
SACM	Structured Assurance Case Meta-Model
SCRA	Supply Chain Risk Assessment
SE	Software Engineering
SHA	System Hazard Analysis
SHARD	Software Hazard Analysis and Resolution in Design
SP	Safety Practitioner
SPP	Safety Practice and its Practitioners
STAMP	System-Theoretic Accident Model & Processes
STPA	System-Theoretic Process Analysis
STRIDE	Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, Elevation of privilege
SSHA	System Safety (or Sub-System) Hazard Analysis
WBA	Why-Because Analysis
ZHA	Zonal Hazard Analysis
ZSA	Zonal Safety Analysis

B List of Abbreviations

See Table 9.