



Assessment of the safety climate in outpatient diagnostic services: Development and psychometric evaluation of a questionnaire

Marianne Jossen¹ · Fabio Valeri² · Christina Heilmaier³ · David Schwappach^{4,5}

Received: 11 May 2018 / Revised: 25 June 2018 / Accepted: 3 July 2018 / Published online: 7 August 2018
© European Society of Radiology 2018

Abstract

Introduction Safe practice and safety culture are important issues in outpatient diagnostic imaging services. As questionnaires assessing safety culture through the measurement of safety climate in this setting are not yet available, the present study aimed to develop and validate such an instrument.

Materials and methods After adaptation of an existing questionnaire and qualitative pretesting, the instrument was tested by collaborators from three outpatient imaging services in Switzerland. Results were first assessed using descriptive statistics. Scores of individual services were compared using a Wilcoxon test assessing differences between rank distributions. The final instrument was tested for validity using inter-rater agreement measures, such as reliability within groups (r_{WG}), and an intraclass correlation coefficient measure (ICC(1)). These measures allowed the assessment of validity of aggregation into a total score ($r_{WG(j)}$) and validated the instrument for its capacity to distinguish various safety climates of different groups by comparing inter-rater agreement in the overall sample to inter-rater agreement of individual services (r_{WG}) and by measuring group effects (ICC(1)). Furthermore, the final instrument was tested for internal consistency and reliability using Cronbach's Alpha.

Results Safety climate scores vary significantly between services. Inter-rater agreement measures show that item aggregation is justified and that the instrument distinguishes various patterns of safety climate. The final instrument proves to be valid, consistent and reliable.

Conclusions The final instrument presents a valid, consistent and reliable option to measure safety climate in outpatient diagnostic imaging services. Results can be used as a basis for quality improvement.

Key Points

- An adapted questionnaire that assesses safety climate in outpatient diagnostic imaging services was developed and tested in Switzerland.
- Psychometric evaluation showed the questionnaire to be a valid, consistent and reliable instrument.
- Results are of interest for imaging services as well as for stakeholders interested more globally in monitoring and quality improvement.

Keywords Outpatient service · Safety · Psychometrics · Organisational culture · Surveys and questionnaires

Abbreviations

aQUA Institut Institut für
angewandte Qualitätsförderung

CSS Culture of Safety Survey
EQUAM Externe Qualitätssicherung
in der ambulanten Medizin

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00330-018-5646-1>) contains supplementary material, which is available to authorized users.

✉ Marianne Jossen
marianne.jossen@icloud.com

¹ EQUAM Foundation, Effingerstrasse 25, 3008 Bern, Switzerland

² Institute of Primary Care, University of Zurich and University Hospital of Zurich, Pestalozzistrasse 24, 8091 Zurich, Switzerland

³ Department of Diagnostic and Interventional Radiology, University Hospital Zurich, Raemistrasse 100, 8091 Zurich, Switzerland

⁴ Swiss Patient Safety Foundation, Asylstrasse 77, 8032 Zurich, Switzerland

⁵ Institute for Social and Preventive Medicine (ISPM), University Bern, Finkenhubelweg 11, 3012 Bern, Switzerland

HSOPS	Hospital Survey on Patient Safety
ICC(1)	Intraclass correlation coefficient
IRA	Inter-rater agreement
r_{wg}	Reliability within groups
$r_{wg(j)}$	Reliability within groups for multiple judgments
SAQ	Safety Attitude Questionnaire

Introduction

As the Bonn Call-for-Action shows, safe practice and safety culture are important issues in radiology. Still, safe practice might be under pressure for various reasons, for example, when competition between services develops, as is the case in Switzerland. In parallel with the healthcare sector in general, the use of diagnostic radiology services has grown in Switzerland over the past several years. From 2010 to 2016, the number of board-certified radiologists rose from 647 to 856 (+32%), while the number of inhabitants per radiologist dropped from 11,593 to 9,836 (-15%) [1] and profits were reduced for a selection of diagnostic imaging procedures [2]. Thus, each service has tried to maximise the operating grade of their infrastructure as much as possible. Consequently, the imaging services staff might work longer hours, run shorter imaging protocols, and/or reduce manpower in order to maintain the services at their current return on investments. However, this might have a negative impact on safe practice and safety culture.

As prior field observations conducted by one of the authors, but also literature suggests [3], safe practice and safety culture in outpatient radiology is not only influenced by radiographers and radiologists but also, to an important extent, by patient administration staff. Indeed, staff administering referrals, planning slots for examinations, scheduling patient appointments as well as instructing patients on preparation for the examination, have an important role in safe practice. To a certain degree, this is similar to the telephone triage performed in physicians' outpatient offices [4]. However, in contrast to general medical offices, members of the patient administrative staff in out-patient diagnostic imaging services often do not have any medical education.

Assessing and comparing safe practice and safety culture across services in a way that respects specific safety issues is gaining more and more importance for services themselves but also for shareholders and initiatives such as the Bonn Call-for-Action that are interested in more global monitoring and quality improvement. One way to assess safety culture is to measure safety climate as perceived by the staff. Safety climate may be defined as shared perceptions or attitudes about the norms, policies, and procedures related to safe practice and patient safety among members of a group, and points to the underlying concept of safety culture, measuring its surface, so to speak [5, 6]. Safety climate is thus associated with

various aspects of safety culture itself having an impact on patient safety. Some aspects of a safety climate can be quite directly related to issues of patient safety. Excessive workload, for example, can lead to taking dangerous shortcuts while performing examinations. For other aspects, the relationship can be somewhat more mediated. Thus, mutual support within the team can be said to foster patient safety by creating a culture of trust where, for example, a lack of knowledge about a certain procedure can be discussed. Thus, studies show that for various instruments, positive associations between safety climate and patient safety can be found [7]. Still, an instrument measuring safety climate specifically in outpatient diagnostic imaging services is, to the best of our knowledge, not yet available.

The measurement of safety climate by questionnaire was first developed for inpatient settings [8–10]. Some of the instruments, like the Hospital Survey on Patient Safety (HSOPS) [11], the Safety Attitude Questionnaire (SAQ) [12] and the Culture of Safety Survey (CSS) [13], have been adapted to a German or Swiss context [14–16]. However, these instruments demonstrate several weaknesses when it comes to outpatient settings [17]. First, outpatient radiology services are often considerably smaller than inpatient settings. Thus, questions related to management or the relationship between departments are difficult to answer. Second, the patient population in outpatient services differ significantly, for example in terms of morbidity, compared to inpatient settings. Safety issues might thus be quite different. Third, these questionnaires strongly focus on medical staff, largely ignoring the important role of administrative personnel without medical qualifications.

The latter point especially often remains an issue when instruments are adapted to outpatient settings and even when they are directly developed for outpatient settings [18, 19]. For example, Modak et al [19] report that radiographers, referral coordinators or outpatient administrative representatives did not answer many items of the SAQ adapted to outpatient settings. Instruments specifically developed for imaging services focus on radiology-specific issues such as magnetic resonance imaging [20] and, again, largely exclude patient administration staff. The SafeQuest, however, which was developed by de Wet et al [21], is designed to suit both medical and administrative staff issues and was specifically created for outpatient services. This instrument uses items issued from 13 existing questionnaires and also incorporates parts of the outpatient version of the SAQ. In their validation study, 35% of respondents classified themselves as being administrative staff [21]. Having been recently translated into German by the aQua Institute (based in Göttingen), this survey was chosen as the basis for this study and adapted to the needs of Swiss outpatient diagnostic imaging services.

The present article describes the adaptation and validation of the SafeQuest survey instrument for outpatient diagnostic

imaging services. The aim was to develop an instrument that is able to assess and compare safety climates across services in a valid, consistent and reliable way.

Materials and methods

Questionnaire development

The designing of the questionnaire was part of a broader initiative that aimed to develop quality indicators for outpatient imaging services. The initiative and the development of the questionnaire was organised by the EQUAM foundation (based in Bern).

First, the German version of the SafeQuest was compared to its English original and some adaptations for the Swiss context were made. No questions about safety concerning specific examinations were added. Further, particular attention was given to only include questions that were equally important for both smaller and bigger organisations. With this choice, the instrument might, on the one hand, miss certain safety climate aspects that mostly concern a specific professional group or questions about specific procedures that might be more common in a certain type of organisation. On the other hand, this choice allowed presenting the same questionnaire to the entire staff while at the same time minimising the risk of low response rates by certain types of staff. After adaptation, a qualitative pre-test [22] was conducted with two radiologists, two radiographers and two people working in patient administration, and was adapted accordingly. Thereafter, the instrument was applied to a sample group for the first time. Survey items are presented as statements asking for the degree of agreement on a seven-point Likert scale ("To what extent do you agree to the following statements?" ranging from 1 – "not at all" to a maximum of 7 – "to a very great extent"). The questionnaire was composed of 30 items and was organised using the thematic structure of de Wet et al [21] thematic structure (see Table 1 for a list of all items). In addition to that, respondents were asked to provide demographic data, i.e. gender, profession, number of years of professional experience and number of years of employment at the respective service. A survey copy is available as [Electronic Supplementary Material](#).

Sample and procedures

The questionnaire was presented as an online survey and invitations for participation were sent to 124 individuals, including employees as well as (co)owners of the services, working at three outpatient imaging services in the German-speaking part of Switzerland. Completion of the questionnaire could be interrupted and continued at another time, but recipients were asked to complete the questionnaire within 2 weeks.

Table 1 Summary of the study sample's characteristics (n=103)

Characteristic	N (%)
Imaging service	
Imaging Service 1	18
Imaging Service 2	21
Imaging Service 3	64
Gender (5 answers missing)	
Male	18 (17)
Female	80 (78)
Profession (3 answers missing)	
Radiologist	16 (16)
Radiographer	51 (50)
Patient administration	33 (32)
Years of working experience (1 missing)	
Less than a year	4 (4)
1–5 years	13 (13)
6–10 years	13 (13)
11–20 years	37 (36)
More than 21 years	35 (34)
Years in the service	
Less than a year	14 (14)
1–5 years	44 (43)
6–10 years	22 (21)
11–20 years	16 (16)
More than 21 years	7 (7)

Participation was voluntary and anonymous. Recipients who had not answered within a 2-week period received a reminder.

Statistical methods

Descriptive statistics (distributions, means, missing answers) assessed the quality and distribution of data. Following Gai et al [23], an individual mean score was calculated for each respondent by aggregating all of his/her questionnaire's item ratings (individual means). Total mean scores and their corresponding distributions were calculated as the overall mean scores of individual mean scores (averaged individual means) and the overall distribution of mean scores. Consequently, the total scores give more weight to bigger organisations in this procedure, but it allows the linking of descriptive statistics to validation measures of content validity, requiring total numbers not be aggregated on an organisational level (see below). Further, the percentage of positive responses, calculated item-wise as recommended by the Agency for Healthcare Research and Quality [24], was calculated. Negatively worded items were reversely coded to ensure that higher scores indicated a more positive assessment of safety climate for every item.

Differences between services were assessed by using a two-sided, two-sample Wilcoxon test, also known as the Mann-Whitney test [25, 26], applied to the distribution of the individual mean scores of the questionnaire. This test assesses differences between rank distributions. A test result < 0.05 was regarded as statistically significant.

Content validity was assessed using measures of inter-rater agreement (IRA). IRA refers to the consensus in scores furnished by multiple judges for one or more targets. In our case, the targets of interest were the total and the item results of the questionnaire, while the judges refer to all persons having answered the questionnaire. The question that IRA measures try to answer is whether scores given by judges show similar patterns, or whether varying judgements appear between individuals or groups of individuals [27].

IRA is most commonly measured by so-called reliability within-group (r_{WG}) indices. The r_{WG} indices compare observed variance in ratings by judges of a certain group to the variance one would expect with random answers and thus measure inter-rater agreement [28–30]. The r_{WG} indices range between 0 and 1, with higher values indicating stronger agreement. For our analysis, we used two different r_{WG} measures: r_{WG} and $r_{WG(j)}$.

As Ginsburg et al [30] describe, $r_{WG(j)}$ measures agreement amongst all judgements (j) – in our case, all the items of the questionnaire. It is a good indicator to assess whether an instrument measures one circumscribable concept and thus item aggregation of total scores is valid. Smith-Crowe et al [31] give significance levels for $r_{WG(j)}$ for questionnaires up to ten items. For instruments with ten items, 100 respondents and seven categories, which is the closest to our present instrument, an $r_{WG(j)}$ of 0.63 is claimed to be sufficient for aggregation of a total score.

An r_{WG} , on the other hand, measures inter-rater agreement on the level of single items – in our case, the total scores of the questionnaire. This measure was used to compare inter-rater agreement specific for each service to inter-rater agreement among all respondents. The hypothesis was that if our instrument proved to be valid, it should be able to distinguish specific safety climates of various services. Thus, r_{WG} among all respondents should be lower compared to that measured for each service. Further, the comparison of r_{WG} across services allows not only the comparison of safety climate levels, but also of safety climate strength [30, 32]. Indeed, the r_{WG} as a measure of IRA shows to what extent the perception of a certain safety climate level, as typically measured by the mean, is strongly present in a service.

Reliability within group measures are well suited to assess and compare patterns of team safety climate for small teams as about 100 respondents are considered to represent large groups [31].

Additionally, the validity of the instrument was measured using the Intraclass correlation ICC(1) measure.

This measure estimates the effect that a certain target, in this case the fact of being a collaborator within a certain service, has upon the ratings of respondents [27, 30]. According to Ginsburg et al [30], an $ICC(1) > 0.05$ shows a group effect and thus confirms that an instrument measures the safety climate perceived by a certain group of individuals.

Internal consistency and reliability were measured with Cronbach's Alpha [33]. This measure can be viewed as the expected correlation of two tests measuring the same construct, varying between 0 and 1. A value of > 0.7 was regarded as sufficient.

All analyses were performed with the Open Source Software R, Version 3.4.3, from 2017 [34].

Results

In total, 106 questionnaires were returned, resulting in a response rate of 86% (106/124). Three questionnaires were excluded due to an excess of missing data, yielding a final sample of 103 completed surveys (83%). Among those persons who answered at least 80% of the questions (99/103) and completed the survey the same day (89/103), the median completion time was 12 min, with a 25%-quantile of 8 min and a 75%-quantile of 21 min. Table 2 summarises the study sample's basic characteristics. As can be seen, the sample includes one larger and two smaller outpatient radiology services. Most of the respondents were women with more than 10 years of working experience, who had been working in the current out-patient service for between 1 and 5 years. Most respondents identified themselves either as radiographers or as working in the service's patient management department (Table 2).

Descriptive statistics, aggregation and safety climate levels and strength

In total, 87% (90/103) of the questionnaires had all safety climate items answered. None of the returned questionnaires presented more than five missing answers for the questions concerning safety climate. Analysis of distributions also showed coherent answering behaviour for reversed items. Therefore, all questions are included in the final instrument and analysis.

Table 1 summarises the results of the final instrument and provides both total scores as well as individual scores of the services. Distribution of answers was close to a normal distribution, as the medians for the total score show, reported in Table 1. Thus, means and standard deviations were used in order to allow for a more detailed view of the discontinuous data of the Likert scale. Analysis for the entire sample revealed the highest score for the statement: "The quality and

Table 2 Results (mean and standard deviation (SD)) of the 30 items that were included in the final instrument. Responses ranged from 1 to 7 on a Likert scale

Item	Total		Service 1 (n=18)		Service 2 (n=21)		Service 3 (n=64)	
	Mean (median)	SD	Mean (median)	SD	Mean (median)	SD	Mean (median)	SD
The performance of collaborators is impaired by excessive workload*	5.0	1.6	6.4	1.0	5.7	1.1	4.3	1.5
Collaborators always have enough time to complete work tasks safely	4.2	1.7	5.6	1.6	4.7	1.4	3.6	1.5
The level of staffing in the service is sufficient to manage the workload safely	4.4	1.8	5.8	1.5	5.1	1.6	3.7	1.7
When pressure builds up, collaborators are expected to work faster even if it means working less accurately*	5.1	1.6	5.9	1.7	5.5	1.6	4.8	1.6
Collaborators feel free to question the decisions of those with more authority	4.4	1.7	5.1	1.4	5.2	1.6	3.9	1.7
Collaborators are comfortable in expressing concerns about the way things are done to the service leadership	4.9	1.8	6.2	1.8	5.3	1.7	4.4	1.7
There is open communication between collaborators, their position notwithstanding	4.6	1.8	5.9	1.5	4.9	1.8	4.1	1.7
Collaborators are kept up to date about service developments	4.4	1.7	5.9	1.6	3.7	1.5	4.3	1.7
The service leadership communicates its long-term plans for the development of the service	4.2	1.9	5.7	1.4	3.8	1.6	3.8	2.0
The hierarchy in the service is a barrier to effective work*	5.6	1.6	6.4	1.8	6.3	1.3	5.1	1.7
Highlighting a significant event will likely result in negative repercussions for the person raising it*	5.8	1.4	6.6	1.5	6.1	1.3	5.5	1.5
The service leadership does not deal effectively with problem collaborators*	5.1	1.7	5.9	1.8	4.9	2.0	4.9	1.5
When collaborators suggest ways to improve processes, the service leadership does not take this seriously*	5.1	1.8	6.4	1.1	5.3	1.6	4.7	1.8
There is a low level of trust between practice team members*	5.7	1.6	6.5	1.8	6.0	1.7	5.4	1.7
Collaborators frequently disregard rules, protocols and procedures*	5.9	1.2	6.4	1.0	6.0	1.3	5.7	1.3
Collaborators treat each other with respect	5.8	1.3	6.3	1.7	5.3	1.6	5.8	1.1
Collaborators always support one another	5.7	1.2	6.3	1.4	5.3	1.5	5.6	1.1
Disagreements within the service team are resolved appropriately	4.8	1.5	6.0	1.9	4.6	1.7	4.5	1.5
Collaborators work well together at all positions within the practice	5.0	1.4	5.7	1.0	5.0	1.1	4.8	1.2
The service is a good place to work	5.9	1.2	6.3	1.6	6.2	1.0	5.7	1.1
Collaborators are generally satisfied with their jobs	5.2	1.3	6.0	0.6	5.6	0.9	5.1	1.3
Good teamwork is promoted and considered to be important by the service leadership	5.2	1.6	6.2	1.4	5.2	1.5	4.9	1.7
All collaborators are encouraged to highlight critical incidents that happen in the service	4.7	1.7	6.2	1.6	4.4	1.5	4.3	1.7
Processes help to prevent critical incidents from happening	4.9	1.4	5.9	1.7	4.5	1.2	4.7	1.4
The opinions of all concerned collaborators are taken into account for the development of processes	4.2	1.8	6.0	1.3	4.4	1.5	3.7	1.7
When processes are changing, the services take time to assess risks for patients, collaborators and the service in advance	4.7	1.7	5.9	1.8	4.8	1.5	4.3	1.7
All concerned collaborators have the opportunity to participate in the analysis of critical incidents	4.3	1.8	6.1	0.9	4.4	1.8	3.8	1.7
The quality and safety of patient care in the service is taken seriously	6.0	1.2	6.6	1.6	6.1	1.2	5.7	1.3
The practice supports the continuing educational development of all collaborators	5.3	1.6	6.2	0.6	5.1	1.6	5.1	1.6
The service encourages learning from the ideas and constructive critique of collaborators in all positions	4.8	1.6	6.1	1.2	4.7	1.5	4.5	1.6
Total score	5.0 (5.1)	1.1	6.1 (6.4)	0.8	5.1 (5.4)	1.0	4.7 (4.7)	1.1

*Reverse coded

safety of patient care in the service is taken seriously”, with a mean of 6.0 on a seven-point Likert scale. In individual analysis within the three services, this statement ranged between

the top to the third highest score. The statement: “The service is a good place to work” received the second highest score (5.9), based on answers from all services. Among the three

services, this statement ranged second for services 2 and 3, and eighth for service 1. The second highest score (5.9) was also obtained for disagreement with the statement: “Collaborators frequently disregard rules, protocols and procedures”. This item scored fourth for service 1, sixth for service 2, and fourth for service 3. Comparing the overall results to positive response rates, the latter were very high for the items “The quality and safety of patient care in the service is taken seriously”, (84%), “Collaborators frequently disregard rules, protocols and procedures”, (84%) and “The service is a good place to work” (85%) with only the item “Collaborators treat each other with respect” scoring higher with 86% positive responses.

In contrast to that, the statements “The service leadership communicates its long-term plans for the development of the service”, “Collaborators always have enough time to complete work tasks safely” and “The opinions of all concerned collaborators are taken into account for the development of processes” scored lowest, with an overall result of 4.2 on a seven-point Likert scale, and also scored lowest for the positive response rates, with 46% for the item asking about participation in development processes and 48% for the two other items. These items ranged between the lowest four scores for the individual services, excluding the second item for service 2 and the third item for service 3 (Table 1). Concerning the positive response rate, the item “All concerned collaborators have the opportunity to participate in the analysis of critical incidents” also scored second lowest with 48%, while it scored a 4.3 as the mean on the seven-point Likert scale.

The $r_{WG(j)}$ for the questionnaire was 0.94 for the 90 completely filled in questionnaires, thus allowing for the aggregation of total scores.

The total mean score for the overall sample was 5.0. Between the outpatient services, the total mean scores ranged from 6.1 for service 1 to 5.1 for service 2, and 4.7 for service 3. When calculating positive response rates, service 3 scored 56% across all items, while service 1 had a positive response rate of 85% and service 2 a positive response rate of 64%.

The two-sided Wilcoxon test for the difference between rank distributions was significant for differences between outpatient services 1 and 2 with a p -value of 0.0003 and between services 1 and 3 with a p -value of < 0.001. The test showed no significant difference between services 2 and 3, with a p -value of 0.12. Figure 1 shows the distributions of individual mean total scores for each service.

Validity, internal consistency and reliability

The r_{WG} ranged between 0.83 for service 1, 0.77 for service 2 and 0.71 for service 3, while the overall r_{WG} was 0.68 and thus smaller than the r_{WG} for each individual service.

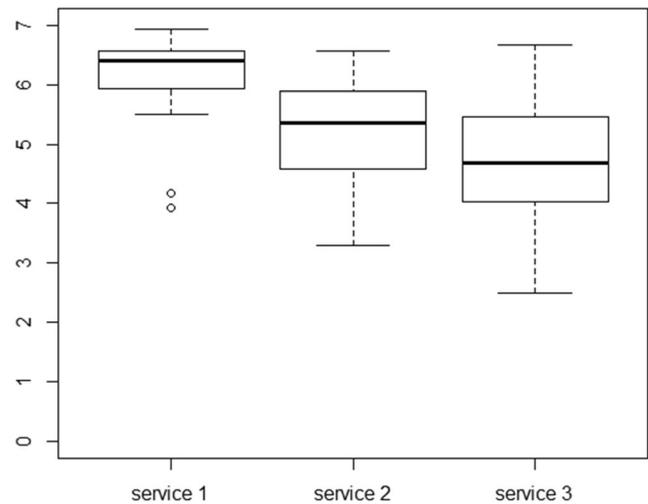


Fig. 1 Comparison of distributions of mean scores of the three services. Responses ranged from 1 to 7 on a Likert scale

The ICC(1) was 0.27, therefore pointing to a group effect. Cronbach’s Alpha was 0.97 with a 95% confidence interval ranging from 0.96 to 0.98.

Discussion

The present study aimed to develop a valid, consistent and reliable questionnaire assessing safety climate in outpatient imaging services.

The response rate was high and the rate of missing answers was very low. Given that the survey was voluntary and anonymous, these results suggest that the statements are easy to understand and relevant to respondents. Important to note, the great proportion of staff working in the call centre and in administrative functions responding to the questionnaire underlines that the instrument is adapted to this professional group, thus avoiding the weaknesses of other instruments [17, 19]. According to literature, our sample involves a sufficient number of participants concerning the measurement of inter-rater agreement [29], but the possibility of comparison of safety climate level between services by means of the Wilcoxon test remains partially limited due to the small size of the participating organisations.

Statements that scored high both in the overall mean scores and positive response rates, as well as in mean scores of the individual services, indicate high safety climate levels for certain elements. Together, the highest-ranking items rather describe general impressions on the priority of safe practice, patient safety and quality as well as the workplace quality and culture.

On the contrary, lower scores were achieved regarding employees’ involvement in the service’s long-term development and improvement of processes, as well as participation in the analysis of critical incidents. Also, time management is an

issue in the participating outpatient imaging services. This might indicate a mounting pressure upon services' staff with regard to increasing competition between different imaging services and monetary constraints after revision of the outpatient reimbursement system. Moreover, to closely monitor the evolution of employees' involvement in processes but also the pressure under which staff is working could be of great interest; not only for the services themselves but also for stakeholders, who are generally interested in quality improvement.

The high value for the inter-rater agreement measure $r_{WG(j)}$ shows that the instrument is indeed measuring a consistent concept, namely safety climate perceptions. On this basis, total scores can be calculated and compared between services or groups of staff.

Our results show differences in total mean scores and significantly varying overall distributions between the three services. Moreover, a higher mean score and distribution corresponded with a higher positive response rate and a higher score for the inter-rater agreement measure r_{WG} . According to the literature [30, 32], it can thus be stated that service 1 incorporates a comparatively high level in a strong safety climate, while service 2 shows an intermediate and service 3 a relatively low level in strength of safety climate. For radiology services, such results can be of interest for further analysis. For instance, a low inter-rater agreement can lead to closer examination of potential differences between professional groups or employees with more or less working experience. High inter-rater agreements, on the other hand, indicate a consistent agreement on a certain climate level. Interventions aiming to improve safety culture can be based on such evidence and can be specifically tailored to certain groups and/or issues.

Comparison between inter-rater agreement of each outpatient service versus inter-rater agreement of the overall sample using the r_{WG} measure reveals that the present instrument is able to measure differing safety climates of individual services. The intraclass correlation measure further strengthens this idea by showing a correlation between the results and the membership to a certain service's team. The instrument thus allows services and regulators to monitor effects of organisational change, for example on staffing or upon safety climate.

Cronbach's Alpha proved to be at a sufficient level, confirming that the final survey instrument is reliable and consistent.

To the best of our knowledge, this is the first validated questionnaire assessing the safety climate of outpatient radiology services as judged by their employees, including personnel working in the call centre or administration. However, a certain selection bias could be present as services participating in the study were highly motivated and already participating in a larger quality improvement project. Comparison of safety climate levels between services remains partially limited due to the small size of certain organisations, pointing to

certain limitations in the quantitative measurement of safety climate in outpatient settings structured in quite small units. Further, it must be said that with its overall sample size, the results of our study cannot claim to be generally applicable to the climate safety levels of outpatient radiology services in Switzerland or beyond. Still, the instrument offers a valid means to conduct more encompassing studies.

The present questionnaire enables the evaluation of the safety climate of outpatient radiology services as perceived by their employees. Results can be used as a basis for quality improvement as well as to monitor the future development of the safety climate in the highly dynamic field of outpatient radiology services.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Marianne Jossen

Conflict of interest The authors of this manuscript declare relationships with the following companies:

Marianne Jossen works for the EQUAM foundation, which uses the developed questionnaires as products.

Statistics and biometry One of the authors, Fabio Valeri, has significant statistical expertise.

Informed consent Written informed consent was not required for this study because data collection was anonymous, and participation was voluntary and did not include medical data.

Ethical approval Institutional Review Board approval was not required because data collection was anonymous, and participation was voluntary and did not include medical data.

Methodology

- Cross-sectional study
- Multicentre study

References

1. Foederatio Medicorum Helveticorum FMH (2017) *Ärztstatistik 2016*. Foederatio Medicorum Helveticorum, Bern. Available at <http://aerztstatistik.myfmh2.fmh.ch>. Accessed 13 march 2018
2. Bundesamt für Gesundheit BAG (2017) [Change of the Ordinance on the Determination and Adaptation of Structures of the Reimbursement System of the Health Insurance]. Bundesamt für Gesundheit, Bern
3. George R, Lau L, Ng K (2014) Team approach to optimize radiology techniques. In: Lau L, Ng K (eds) *Radiological safety and quality. Paradigms in leadership and innovation*. Springer, Dordrecht
4. Gehring K, Schwappach D, Battaglia M et al (2012) Frequency of harm associated with primary care safety incidents. *Am J Manag Care* 19:323–337
5. Weaver SJ, Lubomksi LH, Wilson RF et al (2013) Promoting a culture of safety as a patient safety strategy. A systematic review. *Ann Intern Med* 158:369–374

6. Zohar D, Livine Y, Tenne-Gazit O, Admi H, Donchin Y (2007) Healthcare climate: A framework for measuring and improving patient safety. *Crit Care Med* 35:1312–1317
7. DiCuccio MH (2015) The relationship between patient safety culture and patient outcomes: A systematic review. *J Patient Saf* 11: 135–142
8. Azevedo KB, da Silva CA, Abrantes A (2017) Assessing the hospital survey of patient safety culture. A review of the assumptions, methods and data analysis protocol. *Desenvolvimento e Sociedade* 3:7-31
9. Colla JB, Bracken AC, Kinney LM, Weeks WB (2005) Measuring patient safety climate: A review of surveys. *Qual Saf Health Care* 14:364–366
10. Manser T, Brösterhaus M, Hammer A (2014) You can't improve what you don't measure: Safety climate measures available in the German-speaking countries to support safety culture development in healthcare. *Z Evid Fortbild Qual Gesundheitsw* 114:58–71
11. Sorra JS, Dyert N (2010) Multilevel psychometric properties of the AHRQ hospital survey on patient safety culture. *BMC Health Serv Res* 10:199
12. Sexton JB, Helmreich RL, Neilands TB et al (2006) The safety attitudes questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res* 6:44
13. Weingart SN, Farbstein K, Davis RB, Phillips RS (2004) Using a multihospital survey to examine the safety culture. *Jt Comm J Qual Saf* 30:125–132
14. Pfeiffer Y, Manser T (2010) Development of the german version of the hospital survey on patient safety culture: Dimensionality and psychometric properties. *Saf Sci* 48:1452–1462
15. Zimmermann N, Küng K, Sereika SM, Engberg S, Sexton B, Schwendimann R (2013) Assessing the safety attitudes questionnaire (SAQ), German language version in Swiss university hospitals - a validation study. *BMC Health Serv Res* 13:347
16. Gehring K, Mascherek A, Bezzola P, Schwappach D (2015) Safety climate in Swiss hospital units: Swiss version of the safety climate survey. *J Eval Clin Pract* 21:332–338
17. Schutz A, Counte MA, Meurer S (2007) Development of a patient safety culture measurement tool for ambulatory health care settings: Analysis of content validity. *Health Care Manag Sci* 10:139–149
18. Hoffmann B, Domanska OM, Müller V, Gerlach FM (2009) Development of a questionnaire on the safety climate in family doctor's services (FraSiK): Transcultural adaptation – report on methods. *Z Evid Fortbild Qual Gesundheitsw* 103:521–529
19. Modak I, Helmreich RL, Thomas EJ, Sexton BJ, Lux TR (2007) Measuring safety culture in the ambulatory setting: The safety attitudes questionnaire - ambulatory version. *J Gen Intern Med* 22:1–5
20. Ferris N, Kavnoudias H, Thiel C, Stuckey S (2007) The 2005 Australian MRI safety survey. *AJR Am J Roentgenol* 188:1388–1394
21. De Wet C, Spence W, Mash R, Johnson P, Bowie P (2010) The development and psychometric evaluation of a safety climate measure for primary care. *Qual Saf Health Care* 19:578–854
22. Häder M (2006) *Empiric social research – An introduction*. Verlag für Sozialforschung, Wiesbaden
23. Giai J, Boussat B, Occelli P et al (2017) Hospital survey on patient safety culture (HSOPS): variability of scoring strategies. *Int J Qual Healthcare* 29:685–692
24. Sorra J, Gray L, Streagle S et al (2016) *AHRQ Hospital Survey on Patient Safety Culture: User's Guide*. AHRQ Publication No. 15-0049-EF. Rockville, MD: Agency for Healthcare Research and Quality
25. Bauer DF (1972) Constructing confidence sets using rank statistics. *J Am Stat Assoc* 67:687–690
26. Hollander M, Douglas AW (1973) *Nonparametric statistical methods*. Wiley, New York
27. LeBreton JM, Senter JL (2008) Answers to 20 questions about inter-rater reliability and inter-rater agreement. *Organizational Research Methods* 11:815–852
28. James LR, Demaree RG, Wolf G (1984) Estimating within-group inter-rater reliability with and without response bias. *J Appl Psychol* 69:85–98
29. James LR, Demaree RG, Wolf G (1993) rwg: An assessment of within-group inter-rater agreement. *J Appl Psychol* 78:306–309
30. Ginsburg L, Oore GD (2016) Patient safety climate strength: a concept that requires more attention. *BMJ Qual Saf*. 25:680–687
31. Smith-Crowe K, Burke MJ, Cohen A, Doveh E (2014) Statistical significance criteria for the rWG and average deviation inter-rater agreement indices. *J Appl Psychol* 99:239–261
32. Mascherek AC, Schwappach DLB (2017) Patient safety climate profiles across time: Strength and level of safety climate associated with a quality improvement program in Switzerland - A cross-sectional survey study. *PLoS One* 12:e0181410. <https://doi.org/10.1371/journal.pone.0181410>
33. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334
34. R Development Core Team (2008) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna