Appendix part 1 Example of individual assessments of safety culture in a practice of five health care professionals

Please insert additional figure
**Figure.** Individual assessments of safety culture in a practice of five health care professionals. The figure displays the nine FraTrix dimensions (spokes) and the five culture grades from red (dissmissive) to dark green (generative). Each individual assessment is marked with a blue dot.
Appendix part 2. Method details

Recruitment

A short invitation letter was sent to all 1,629 registered general practices in southern Hesse. Then each physician interested in the trial was contacted by phone and further information was sent to the practice. Participating teams were offered a small honorarium (up to €800 depending on team size) to be received after final data collection. Practices that decided not to participate were contacted after the end of the inclusion phase and their key characteristics assessed (practice model, team size, location, and type of quality management system).

Randomisation

After obtaining consent, practices were randomly assigned to intervention and control in a 1:1 ratio using randomly permuted blocks of four, six, and eight (www.randomization.com). Randomisation was stratified by practice model (single-handed or group practice), with the practice model firstly linked to team size, which we assumed may influence the structure and processes in the practice and therefore also outcome measures, and secondly to the different kind of leadership (single owners or partners), as this may have an impact on leadership's commitment to patient safety and thus on safety culture. Although the required sample size was only 60 practices, we randomised all 65 practices that agreed to participate in case of withdrawals. Randomisation was conducted by a non-member of the project team. The assignment sequence was unknown to the study team and was stored in concealed envelopes until baseline data were collected.

Sample size

The calculation was based on data from 302 practices belonging to the European Practice Assessment system EPA (Institute for Applied Quality Improvement and Research in Health Care (AQUA), Goettingen) as of 2008. The EPA indicator for error management consists of six criteria that are based on self-reported data collected during a practice visit. Meeting the criteria indicates a fully implemented error management system.

Based on the EPA results and the results from our own pilot study, only 2% of control practices were expected to fulfil all six criteria, 22% were expected to fulfil five, 55% four, 12% three, and 5% two criteria. About 3% of practices were expected to fulfil only one criterion, and 1% of the practices were expected to meet none of the criteria. We assumed that FraTrix would be effective and that practices in the intervention group would tend to fulfil more criteria. Thus, in the intervention group, 20% of practices were expected to meet all six criteria, 43% five, and 31% four criteria. Only 3% of practices were expected to fulfil three, 2% two, 1% one, and 0% none of the criteria. This pattern of results corresponds to a

---

a The German primary health care system differs from the English system in several respects: Mandatory gatekeeping does not exist. Patients have direct access to general practitioners as well as to specialists in ambulatory practice. Nevertheless, the majority of the German population regularly consults a general practitioner. In most general practices, one or two physicians work with up to five health care assistants (HCA), who do administrative and clerical work as well as basic clinical procedures. The physicians usually own their surgeries.

competing probability (CP)\(^c\)\(^d\) in the Wilcoxon-Mann-Whitney test of 0.73, with CP values greater than 0.50 indicating greater efficacy in the experimental group. To detect a CP of 0.73 with a power of 80% and a two-sided significance level of 5%, a total of 45 practices will be needed.

Data management

Two blinded raters independently judged whether the criteria for given indicators were fulfilled. Consensus on the criteria had to be achieved. Since the indicators consisted of different numbers of criteria, we standardised the scores by dividing the sum of fulfilled criteria by the total number of criteria per indicator. For the FraSiK questionnaire, scale scores on an individual level were calculated when a participant responded to at least 75% of the items relating to a specific scale. Scale scores were calculated as the mean of the scale’s items. Safety climate scores on a practice level were calculated as the mean of the respective team members’ scores.

Quality of incident reports and improvement protocols (from intervention group practices only) were evaluated independently by two blinded raters who again had to find a consensus in case of disagreement. In order to assess the quality of assessments on safety culture indicators, quality of incident reports and improvement protocols, we assessed inter-rater-agreement by calculating Cohen’s Kappa \(\kappa\).\(^e\) The duration of the trial was 12 months between baseline and final data collection. No outcome measure was changed during the trial.

Quality of data

Inter-rater-agreement on the assessment of PSCI criteria proved to be substantial (Cohen’s \(\kappa\) from 0.6 to 0.8) for 15.4% of ratings; or almost perfect (\(\kappa > 0.8\)) for 78.8% of 104 ratings. With regard to inter-rater-agreement on the assessment of incident reports, 50% of a total of 12 assessments were evaluated as substantial (41.7%) or almost perfect (8.3%), similar to the improvement protocols in the intervention practices (Cohen’s \(\kappa\) from 0.6 to 0.8; 33.3%, >0.8; 20% of 15 ratings in total).

Statistical analysis

Primary analysis followed the intention-to-treat principle with all practices analysed as randomised, and was conducted with baseline observation carried forward for missing values. Continuous data at baseline and at 12-month follow-up were summarised by using means and standard deviations (SD), and categorical data by using frequency counts and percentages. Twelve months after randomisation, we analysed the primary outcome, error management, with the Wilcoxon-Mann-Whitney (WMW) test at a two-sided 0.05 significance level. The effect of the intervention on the primary outcome is expressed as competing

\(^c\) Zhao YD, Rahardja D, Qu Y. Sample size calculation for the Wilcoxon-Mann-Whitney test adjusting for ties. Stat Med 2008;27:462-8


probability (CP) from the WMW test statistic with the corresponding 95% confidence interval (CI). A value of 0.5 for the CP indicates no difference between intervention and control group, a CP greater than 0.5 indicates that practices in the intervention group performed better than those in the control group, and a CP of less than 0.5 indicates that intervention practices performed worse than control practices. Non-parametric repeated measures analysis of variance (ANOVA) was used to conduct a sensitivity analysis. The non-parametric ANOVA model included terms for group (intervention, control), time (baseline, 12 months), and for the stratification variable, practice model (single-handed practice, group). Group-by-time interaction, group-by-practice model interaction, and practice model-by-time interaction were also included.

Secondary endpoints were analysed using the same statistical methods described for the primary endpoint. All tests were two-sided at a significance level 0.05 and were not adjusted for multiple comparisons. Thus, only the result of the primary efficacy analysis was interpreted in a confirmatory manner.

**Ethical issues**

The ethical guidelines for quality improvement methods were all met. Participation of practice teams was completely voluntary. Patients did not take part in the trial and patient data – when used – were anonymised. The institutional review board of the Medical Faculty of the Goethe University Frankfurt stated that no review of ethical issues was required.

This report follows the CONSORT guidelines for non-pharmacologic trials and the SQUIRE publication guidelines for reporting health care quality improvement research.

---

Appendix part 3. Patient safety culture indicators used to assess patient safety culture in the FraTrix trial. Indicators were derived from established quality indicators in general practice (the source of the respective indicators is provided below) For each indicator, criteria for fulfilment were defined. About 100 interview items were created to assess the criteria during data collection visits. Assessment of the criteria was by self-report of an expert in the practice (below marked by S), via a chart review of randomly selected patients (C), or by review of practice documents or other equipment (D), as appropriate. During the trial, data collection was conducted by well-trained project team members who were blinded regarding the allocation.

1. **Basic life support training.** “Basic life support training is carried out in the practice in accordance with specific norms and at least once per year.”

   - Criterion 1 (S): A training course on emergency measures (practical exercises or in another form) is carried out once a year.
   - Criterion 2 (S): The entire practice team participates in the training (physicians and health care assistants).
   - Criterion 3 (S): Basic life support training is carried out (practical exercises – emergency care).

2. **Emergency medications.** “The shelf life of emergency medications in the practice is checked using a standard procedure at least once a year.”

   - Criterion 1 (S): In the practice, there is a list of all emergency medications that must be available at all times.
   - Criterion 2 (S): At least once a year, the shelf lives of emergency medications are inspected to see if they have passed, and replaced if necessary. For this there is a (written or established) directive.
   - Criterion 3 (S): The inspection is carried out according to a standard uninterrupted procedure.
   - Criterion 4 (S): The results of the inspection are documented.
   - Criterion 5 (D): The shelf lives of essential emergency medications have not passed.
   - Criterion 6 (S): The place where emergency medications are kept is known to everyone in the practice team.

3. **Critical incidents.** “Over the last year, critical incidents that occurred in the practice were discussed by the whole team on at least four occasions.”

   - Criterion 1 (S): At least four incidents were discussed.

---


Criterion 2 (S): In the practice, critical incidents are discussed by the entire team.

Criterion 3 (S): The team discussions deal with causes and preventive measures.

Criterion 4 (S): The results of the discussion are documented.

Criterion 5 (S): The results of the discussion are communicated to the entire team.

4. **Error management.** “The practice carries out systematic error management.” *(primary outcome)*

Criterion 1 (D): The practice keeps a reporting system of critical incidents. At least one entry has been made over the last 12 months.

Criterion 2 (S): Critical incidents are analysed by the entire team. At least one critical incident has been analysed by the entire team during the last three months.

Criterion 3 (S): The practice can name a specific example of a critical incident that took place during the last 12 months. The event fulfils the criteria for a patient safety incident.

Criterion 4 (S): Measures are taken in the practice to prevent critical incidents. In the practice, responsibility for the implementation of these measures, and the period over which the implementation is to take place, are also fixed. In addition, the practice learns from incidents that have occurred in other practices and takes measures to avoid critical incidents accordingly.

Criterion 5 (S): The practice can name an example of a specific measure taken over the last 12 months to avoid critical incidents. The measure was also implemented.

Criterion 6 (S): The practice actively participates in an external error reporting system.

5. **Complaints management.** “The practice operates a complaints management system.”

Criterion 1 (D): In the practice, there is a clearly visible slip box for written patient complaints and suggestions.

Criterion 2 (S): The entire practice team discusses patient criticisms and complaints.

Criterion 3 (S): Criticisms and complaints made by patients are documented in this practice.

Criterion 4 (S): The practice can name an example of a patient’s suggestion that was then implemented. One example and the resulting measure are sufficient.

Criterion 5 (S): The practice carries out patient surveys every three years.

6. **Documentation of allergies.** “The documentation of allergies to medications is carried out according to a standard procedure and the information is clearly in evidence.”

Criterion 1 (C): Medication allergies are separately documented in patients’ records.

Criterion 2 (C): Such entries are clearly visible when prescriptions are written.

7. **Medication list.** “An overview of all current medications can be taken from patient records at all times.”
Criterion 1 (C): An overview of all current medication prescriptions (long-term medication) written in this practice can be found in the corresponding patient’s records.

Criterion 2 (C): Every medication plan includes an entry that is not older than 12 months.

Criterion 3 (S): In this practice, patients are asked at least once a year, what over-the-counter medication they take.

Criterion 4 (S): Every medication plan includes information on current medication prescriptions written by other doctors and specialists that are treating the patient. At least once a year, patients are asked what medications they have been prescribed by other doctors.

Criterion 5 (S): The current medication plan is checked whenever a prescription is written.

8. **Monitoring of repeated prescriptions.** “In this practice, there is a method for checking repeat prescriptions and ensuring they are really indicated.”\(^b\)

Criterion 1 (S): An examination of long-term medications occurs at least once a year to ensure they are still indicated.

Criterion 2 (S): Patient medication adherence is checked (the right medication, not taken for too long or too short a period).

Criterion 3 (S): The number of repeat prescriptions written without any contact to the doctor is limited.

9. **Repeat prescribing.** “In this practice, a secure procedure is in place for issuing repeat prescriptions.”\(^b\)

Criterion 1 (S): The printed prescription is checked a second time (by a second person, or by the doctor when he or she signs it) before being given to the patient.

Criterion 2 (S): Long-term medications are clearly indicated as such in the patient’s chart.

Criterion 3 (S): Prescriptions can only be picked up at specific times during the practice’s opening hours.

10. **Oral anticoagulants.** “The practice has processes in place that safeguard oral anticoagulant therapies.”\(^l\)

Criterion 1 (C): There is a clear indication of anticoagulant therapy in the records of every patient that is receiving such treatment.

Criterion 2 (C): International Normalized Ratio (INR) levels are checked at least every four weeks.

---

Criterion 3 (C): A patient's most recent INR value is noted in his or her records.

Criterion 4 (C): A patient's target INR level is noted in his or her records.

Criterion 5 (C): A patient's current oral anticoagulant dosage is noted in his or her records.

11. **Laboratory tests.** “In this practice, there is a secure procedure for carrying out laboratory examinations and following up on the findings.”

Criterion 1 (S): The practice has a method to ensure that in case of laboratory examinations (e.g. blood test), the sample and the patient have been correctly identified.

Criterion 2 (S): The practice documents outgoing requests to other service providers (e.g. samples dispatched to the laboratory).

Criterion 3 (S): Laboratory findings are documented in the practice and matched to dispatched requests, thus ensuring an overview of results that are still outstanding.

Criterion 4 (S): The practice has a procedure in place that ensures that incoming laboratory findings are seen by the consulting physician.

Criterion 5 (S): Laboratory findings are entered into the patient's records.

Criterion 6 (S): The measures taken as a result of incoming laboratory findings are documented in the patient's records.

12. **Flu vaccination.** “Patients over 65 years of age are offered a flu vaccination once a year.”

Criterion 1 (S): All patients over the age of 65 are offered a flu vaccination once a year.

Criterion 2 (S): A recall system is in place for the annual flu vaccination.

Criterion 3 (C): Whether a patient over the age of 65 has received a flu vaccination in the practice or elsewhere is documented in his or her records.

---

Appendix part 4. Quality assessment of patient safety incident reporting adapted from McKay et al.\textsuperscript{21}

Does the report describe a patient safety incident\textsuperscript{20}? If yes, then the following criteria were checked and graded accordingly (grades are given in brackets).

I. Quality of patient safety incident report
   a) Are causes or contributing factors given for the occurrence of the incident? (Yes = 1, no = 0)
   b) Are any consequences of the incident described? (Yes = 1, no = 0)
   c) Are any measures described to prevent the incident or a similar one from happening again? (Yes = 1, no = 0)
   d) Appropriateness of this/these measure/s for the prevention of this or similar incidents in the future\textsuperscript{c}
      - Not appropriate at all (e.g., contributing factors or causes of the incident are not taken into account; e.g. “next time be more careful”) (Yes = 0)
      - Appropriate for the prevention of this precise incident (Yes = 1)
      - Appropriate for the prevention of similar incidents (Yes = 1)
      - Appropriate for the prevention of different kinds of incident (e.g., measures that have an impact on latent conditions or relevant contributing factors) (Yes = 1)
   e) Any follow-up on measures taken to prevent the described incident? (Yes = 1, no = 0)

   Quality score for each incident report = sum of criteria a to e (minimum score of ‘0’, maximum score of ‘5’).

II. Number of reported patient safety incidents

III. General management of patient safety incidents

Each report was globally assessed in terms of incident management without regard to the above mentioned quality criteria:

- no incident management at all (0)
- poor incident management (1)
- fair incident management (2)

\textsuperscript{c} If more than one measure is described, the sum of the grades is divided by the number of measures, e.g., three measures, with two of them describing measures appropriate for the prevention of similar incidents would yield a score of $2/3 = 0.67$. 
- good incident management (3)
- excellent incident management (4)
### Appendix part 5. Scores for all patient safety culture indicators at baseline and 12 months after randomisation for both intervention and control groups.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>At baseline</th>
<th>At 12 months</th>
<th>Group comparison at 12 months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interv. (n = 28)</td>
<td>Control (n = 32)</td>
<td>Interv. (n = 28)</td>
</tr>
<tr>
<td>Error management (primary outcome)</td>
<td>0.47 (0.24)</td>
<td>0.56 (0.22)</td>
<td>0.66 (0.18)</td>
</tr>
<tr>
<td>Basic life support training</td>
<td>0.62 (0.37)</td>
<td>0.66 (0.30)</td>
<td>0.45 (0.46)</td>
</tr>
<tr>
<td>Emergency medications</td>
<td>0.65 (0.19)</td>
<td>0.62 (0.19)</td>
<td>0.63 (0.20)</td>
</tr>
<tr>
<td>Critical incidents</td>
<td>0.58 (0.30)</td>
<td>0.71 (0.25)</td>
<td>0.64 (0.24)</td>
</tr>
<tr>
<td>Complaints management</td>
<td>0.56 (0.23)</td>
<td>0.63 (0.18)</td>
<td>0.58 (0.26)</td>
</tr>
<tr>
<td>Documentation of allergies</td>
<td>0.82 (0.28)</td>
<td>0.69 (0.33)</td>
<td>0.79 (0.25)</td>
</tr>
<tr>
<td>Medication list</td>
<td>0.52 (0.27)</td>
<td>0.52 (0.25)</td>
<td>0.56 (0.20)</td>
</tr>
<tr>
<td>Monitoring of repeated prescriptions</td>
<td>0.60 (0.31)</td>
<td>0.60 (0.31)</td>
<td>0.60 (0.28)</td>
</tr>
<tr>
<td>Repeat prescribing</td>
<td>0.50 (0.23)</td>
<td>0.46 (0.26)</td>
<td>0.44 (0.18)</td>
</tr>
<tr>
<td>Oral anticoagulants</td>
<td>0.44 (0.17)</td>
<td>0.42 (0.17)</td>
<td>0.41 (0.19)</td>
</tr>
<tr>
<td>Laboratory tests</td>
<td>0.68 (0.16)</td>
<td>0.72 (0.14)</td>
<td>0.71 (0.15)</td>
</tr>
<tr>
<td>Flu vaccination</td>
<td>0.45 (0.23)</td>
<td>0.46 (0.20)</td>
<td>0.40 (0.23)</td>
</tr>
</tbody>
</table>

Note—Indicator scores were calculated by dividing the sum of fulfilled criteria by the total number of criteria per indicator, resulting in a score between 0 and 1; ‘0’ denotes ‘no criterion achieved’ and 1 ‘all criteria achieved’. Mean and standard deviation are presented for all indicators.