

BMJ Open Improving patient safety through better teamwork: how effective are different methods of simulation debriefing? Protocol for a pragmatic, prospective and randomised study

Julia Freytag,¹ Fabian Stroben,^{2,3} Wolf E Hautz,³ Dorothea Eisenmann,^{2,4} Juliane E Kämmer^{5,6}

To cite: Freytag J, Stroben F, Hautz WE, *et al*. Improving patient safety through better teamwork: how effective are different methods of simulation debriefing? Protocol for a pragmatic, prospective and randomised study. *BMJ Open* 2017;7:e015977. doi:10.1136/bmjopen-2017-015977

► Prepublication history and additional material for this paper are available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-015977>).

JF and FS contributed equally.

Received 16 January 2017

Revised 2 May 2017

Accepted 18 May 2017



CrossMark

For numbered affiliations see end of article.

Correspondence to

Fabian Stroben; fabian.stroben@charite.de

ABSTRACT

Introduction Medical errors have an incidence of 9% and may lead to worse patient outcome. Teamwork training has the capacity to significantly reduce medical errors and therefore improve patient outcome. One common framework for teamwork training is crisis resource management, adapted from aviation and usually trained in simulation settings. Debriefing after simulation is thought to be crucial to learning teamwork-related concepts and behaviours but it remains unclear how best to debrief these aspects. Furthermore, teamwork-training sessions and studies examining education effects on undergraduates are rare. The study aims to evaluate the effects of two teamwork-focused debriefings on team performance after an extensive medical student teamwork training.

Methods and analyses A prospective experimental study has been designed to compare a well-established three-phase debriefing method (gather–analyse–summarise; the *GAS method*) to a newly developed and more structured debriefing approach that extends the GAS method with *TeamTAG* (teamwork techniques analysis grid). TeamTAG is a cognitive aid listing preselected teamwork principles and descriptions of behavioural anchors that serve as observable patterns of teamwork and is supposed to help structure teamwork-focused debriefing. Both debriefing methods will be tested during an emergency room teamwork-training simulation comprising six emergency medicine cases faced by 35 final-year medical students in teams of five. Teams will be randomised into the two debriefing conditions. Team performance during simulation and the number of principles discussed during debriefing will be evaluated. Learning opportunities, helpfulness and feasibility will be rated by participants and instructors. Analyses will include descriptive, inferential and explorative statistics.

Ethics and dissemination The study protocol was approved by the institutional office for data protection and the ethics committee of Charité Medical School Berlin and registered under EA2/172/16. All students will participate voluntarily and will sign an informed consent after receiving written and oral information about the study. Results will be published.

Strengths and limitations of this study

- The study design builds on established principles of teaching and assessing teamwork.
- The study will be one of the first to explore the effects of teamwork-focused debriefing on team performance with undergraduate medical students.
- The study will be embedded in a well-established simulation setting with proven efficacy.
- The study will be a pragmatic, randomised comparison of two debriefing methods.
- Only a single centre will be studied.
- Feedback quality will not be externally evaluated.

INTRODUCTION

Medical errors and adverse events occur with an incidence of about 9% and can seriously harm patients.^{1 2} Error rates in emergency settings are even reported to be twice as high.^{3–5} Most medical errors originate from human factors and teamwork⁶ or medication errors⁷ and about half of all medical errors are considered preventable.^{1 7}

Empirical evidence^{6 8–11} suggests that improving teamwork may be key to reducing medical error. Yet, although teamwork and patient safety are prominent objectives in many national outcome frameworks,^{12–14} these topics are insufficiently represented in undergraduate education and are rarely assessed, even though validated teamwork assessment tools exist.^{15 16} Consequently, about 60% of junior doctors in Germany reported feeling inadequately prepared for clinical practice¹⁷ and almost half of the residents in a Canadian survey reported feeling overwhelmed when leading a resuscitation team.¹⁸

In addition, common interventions targeting the quality of teamwork and human

factors, such as simulation training and crisis resource management (CRM) training, have produced a variety of effects.^{19 20} In both simulation and CRM training, debriefing is considered crucial to enhancing learning²¹ but little is known about how best to debrief. In fact, the widely differing effects of simulation may very well result from differences in debriefing. A feasible and beneficial debriefing method, particularly for undergraduates, could lead to more effective simulation sessions and thus ease the transition into clinical practice for junior doctors. This could ultimately lead to a reduction of medical errors and thus improved patient outcome. In this study we will compare the effects of two different debriefing methods on team performance and the acquisition of teamwork skills during teamwork simulations for medical students.

Training and debriefing

The concept of CRM was originally derived from safety training in aviation and has been adapted to the health-care sector, another high-stakes environment.²² The idea of CRM is to guide individuals and teams in emergency situations (crises), encouraging them to use all available resources to manage the situation effectively and prevent critical incidents from occurring in the first place. CRM training has been shown to be a potent tool to improve teamwork and—as a consequence—patient safety.^{23–25} In our study, elements of CRM set the framework for teamwork training and debriefing during an emergency room simulation.

Simulation debriefing is defined as a bidirectional and interactive discussion after a simulation in which participants reflect on their actions and analyse their performance.²¹ Feedback is a central process element of debriefing that is often used as a conversational technique especially in participants with little experience in debriefing.²⁶ Feedback is defined as the delivery of information to improve reasoning or behaviour compared with defined performance standards,^{26 27} and it is critical in improving learning.²¹ How best to integrate feedback into debriefing, what specific aspects to address and how to structure debriefing to foster learning are, however, still unknown.^{21 28} The goal of this study is thus to evaluate the potential benefit of preselecting certain aspects to be discussed during debriefing and of structuring debriefing with the help of a cognitive aid. To this end, we will compare a well-established debriefing method to a more structured and feedback-focused method to evaluate their effects on teamwork, learning opportunities, feasibility and helpfulness for participants (and instructors). We will focus on two debriefing methods, the *gather–analyse–summarise* (GAS) method and the GAS method plus a cognitive aid:

1. The GAS method: This debriefing method consists of three parts: gathering, analysing and summarising.^{29 30} The GAS method is one of many similar three-step debriefing structures²⁶ and has been used, for example, in simulation courses run by the American Heart Association.³⁰ During the first

phase (gather), participants are given the opportunity to report their thoughts on the simulated situation. They are encouraged to exchange their views on what actually happened to establish a shared mental model of the situation. This model can afterwards be used to discuss the simulation in a learner-centred way (analyse). During this process, questions tailored towards specific learning objectives are used to facilitate participants' reflection on and analysis of their actions and induce learning. Finally, the debriefing is summed up and critically reviewed by the team and its instructor (summarise).^{26 29} Topics discussed during the debriefing using this method are mostly self-selected by the team and instructor, which makes this method highly flexible. A possible drawback with regard to teamwork (or any other specific learning objective) is that its potential to enhance the quality of teamwork is influenced by the instructor's level of experience.²⁶ A typical question to start the debriefing with the gather step might be 'How do you feel now?' followed in the analysis step by 'What worked well?' or 'Do you see any opportunities for improvement?' The summarise step might be initiated by 'What we learned from this session....'

2. The GAS method plus a cognitive aid: This newly developed debriefing method uses the GAS structure detailed above and additionally provides the instructors with a cognitive aid to structure the debriefing in more detail. It further provides a selection of important aspects to address during debriefing. Cognitive aids are 'structured pieces of information designed to enhance cognition and adherence to...best practices.'³¹ Cognitive aids have been shown to be beneficial in different areas of medicine.^{32–34} Moreover, cognitive aids are useful for debriefing: Instructors' use of a cognitive aid may improve participants' acquisition of behavioural and cognitive outcomes after simulation—especially so with novice instructors.³⁵ In practice, such aids are often a pocket card, script or poster.

We will use a specific cognitive aid called 'TeamTAG' (teamwork techniques analysis grid) to foster observation and feedback relevant to teamwork. TeamTAG is a guideline for structuring the feedback process during debriefing and remembering what to address during the analysis step of the GAS method. The TeamTAG lists teamwork-relevant CRM principles together with descriptions of behavioural anchors that serve as directly observable patterns of teamwork and provides space for notes (see online supplementary information). The TeamTAG can be printed on a single sheet of paper (A4) and filled in during observation of the simulation. After the simulation, instructors have the flexibility to set priorities for debriefing based on their observations and structured notes. The debriefing itself will follow the same structure as under the GAS method. However, the TeamTAG might, for example, remind instructors that team leaders

'allocate roles & tasks' or are responsible for 'monitoring progress' (according to the CRM principle 'exercise leadership and followership'). These aspects might be specifically addressed by group instructors to improve group reflection during the analysis step.

Hypotheses

First, we assume that the GAS method plus TeamTAG will be a more effective debriefing tool than the common GAS method alone and will lead to the discussion of more teamwork-relevant principles. Debriefing using the GAS method plus TeamTAG should thus result in more learning opportunities for teams and ultimately in improved team performance. This hypothesis is based on the fact that the TeamTAG is concise and guides observation and feedback with practical examples. Using these examples during observation may help focus the observers' attention³⁶ and result in the team discussing more teamwork-relevant CRM principles. In undergraduate education, instructors are often novices and vary considerably regarding how experienced they are in debriefing. Because novices were shown to benefit more from structured debriefing scripts than more experienced instructors,³⁵ we consider our environment (see the Methods and analysis section) ideal for detecting differences between the two debriefing methods if they exist.

Hypothesis 1a: Participants who receive debriefing based on the GAS method plus TeamTAG will show a greater improvement in team performance than those who discuss the simulation according to the common GAS method alone.

Hypothesis 1b: Participants who receive debriefing based on the GAS method plus TeamTAG will report discussing a higher number of CRM principles than participants who are debriefed with the GAS method alone.

Second, we expect that teams receiving debriefing based on the GAS method plus TeamTAG will perceive teamwork skills as more important after the simulation event, which should increase their sensitivity to a culture of safety and the likelihood of changing their behaviour.^{37 38} Moreover, perceiving the content of the debriefing as more important should lead to higher overall satisfaction with and perception of helpfulness of the debriefing.

Hypothesis 2a: Participants who receive debriefing based on the GAS method plus TeamTAG will report a higher level of perceived importance of teamwork principles than those who are debriefed according to the common GAS method.

Hypothesis 2b: Participants who receive debriefing based on the GAS method plus TeamTAG will report higher satisfaction with and helpfulness of the debriefing they received than those who are debriefed according to the GAS method alone.

Third, we will focus on the satisfaction of the instructors as a measure of feasibility and efficiency. We expect higher satisfaction when they use the GAS method plus TeamTAG as it might facilitate more structured feedback

and it provides a better opportunity for instructors to address the learning objectives of their participants.

Hypothesis 3: Instructors who use the GAS method plus TeamTAG will report higher levels of feasibility and efficiency of their debriefing than instructors who use the GAS method alone.

METHODS AND ANALYSIS

This investigation is designed as a prospective experimental superiority study with intervention and control groups receiving debriefing during a simulation training based on either the GAS method plus TeamTAG or the GAS method alone, respectively. The study will be executed during an emergency department (ED) simulation at Charité Medical School, Berlin, Germany, on 14 January 2017. The ED simulation has been implemented at the local skills laboratory since 2013 on a peer-led basis. The main goal of this extensive, 8-hour night-shift simulation training is to give students the opportunity to experience being the person in charge of a patient's healthcare. This event takes place once a year, with about 35 students in their final year of medical studies participating voluntarily. Participants are recruited via newsletter and advertising posters. The students act in randomly assigned teams of five and self-select into different roles (team leader, team member, observer), which they switch during the night. Simulated patients and high-fidelity simulators are used to create realistic case simulations; simulated radiological and laboratory services are provided. One of the main goals of the event is to improve students' confidence in working with medical emergencies in an ED over the course of the night.³⁹ The simulation was awarded a project prize by the German Association for Medical Education in 2016.

Each student team has to work on six simulated cases. Each case is staffed with a case instructor who is responsible for the simulation and provides technical help. Each student team is accompanied by a group instructor who guides the participants during the night. After every case, multisource feedback is provided by simulated patients, observing participants and case instructors. As part of our study, in 2017 participants will additionally receive a teamwork-based debriefing by the group instructors after every case in one of two conditions (GAS method vs GAS method plus TeamTAG). Additionally, the quality of teamwork will be rated by trained raters throughout the night.

As group instructors we will choose experienced peer teachers who are advanced in their healthcare studies (medicine, nursing) and have completed emergency room courses/electives during their studies. Peer teachers at Charité Medical School Berlin frequently give courses in clinical skills training and simulator-based emergency medicine trainings for other medical students. All group instructors undergo extensive feedback training during their studies and are furthermore trained in working with and debriefing groups.

Development of the TeamTAG as cognitive aid

As a basis for this study, the TeamTAG guideline was developed with the goal of having a feasible and time-efficient feedback instrument that supports teaching basic teamwork skills to participants. Two investigators (JF and FS) developed the TeamTAG guidelines that present six common CRM principles,^{22 40} each accompanied by the description of behavioural anchors. The six principles are (1) anticipate and plan ahead, (2) set priorities dynamically, (3) call for help early, (4) exercise leadership and followership, (5) communicate effectively and (6) re-evaluate repeatedly. The TeamTAG can be found in the online supplementary material. The CRM principles and their behavioural anchors were chosen to fit the following criteria: (A) simulation setting, (B) presumed skills of participants, (C) experience of instructors and (D) observability. The tool was reviewed and adjusted by an experienced group of anaesthesiologists, emergency medicine physicians, simulation instructors and peer tutors, all experienced in medical education and simulation-based learning. In a prestudy, feasibility for instructors was examined (see the Preliminary results section) but not compared with an approach without the TeamTAG.

Team performance measurement

To measure team performance, we will use the Team Emergency Assessment Measure (TEAM).¹⁵ TEAM is an assessment tool that has been applied to both clinical and simulation environments.^{15 16 41} It consists of 11 items belonging to the three subscales leadership, teamwork and task management. Example items are 'the team leader maintained a global perspective' and 'the team prioritized tasks', measured on a 5-point Likert scale of 0 (*never*) to 4 (*always*). Additionally, it includes an overall rating of team performance (range: 1 (*very poor performance*) to 10 (*very good performance*)).

As there was no German version of the TEAM, the English version was translated into German using elements of the TRAPD (translation, review, adjudication, pretest, documentation) methodology.⁴² Two investigators (JF and FS) independently translated the TEAM into German in parallel, reviewed the results and consented to one version, which was translated back by a native English speaker. This new version was compared with the original TEAM and agreed to by both investigators and the native speaker. All steps of the translation were documented.

After the TEAM was translated, we developed a rater training. The training involves three aspects that are important in preparation for accurately assessing a certain behaviour or skill^{1 43}: a *rater error training* in which information is provided on typical rating errors to raise awareness and prevent them,² a *performance dimension training* to teach raters about the targeted dimensions, including definitions and videotaped examples, and³ a *frame-of-reference training*, in which videotaped examples showing teamwork of different levels of quality are assessed and discussed. All raters who will be responsible

for TEAM ratings in this study (case instructors and additional raters) will receive this rater training and additional written material on teamwork and how to use the TEAM.

Group instructors debriefing training

Before data collection, all group instructors will receive a teamwork-related training and additional written material with information about how to provide feedback and conduct debriefings and about human factors in general and CRM in particular, which is intended to serve as a framework for discussing all teamwork aspects during debriefing. The training will include videos showing good and bad examples of teamwork and will be followed by discussions about opportunities for debriefing in these specific situations (adapted from *frame of reference training*⁴³). After this training, which will be the same for all group instructors, the instructors will be randomly assigned, stratified by level of academic education and additional professional training (eg, nurse or paramedic), to the two conditions. The two groups will receive separate instruction from the investigators: The intervention group instructors will be told to discuss their groups' performance with the help of the TeamTAG and to focus on each CRM principle of the TeamTAG at least once during the first five cases (ie, one or two principles per case) so that by case 6 all CRM principles will have been debriefed and team performance during case 6 can be compared between conditions. Furthermore, they will be instructed to re-evaluate their previous focus of debriefing after each case if behaviour does not change sufficiently from their perspective. The order of chosen topics can be varied by the instructors and should be adjusted to observed difficulties in teamwork during the simulation. The control group instructors will be advised to give feedback regarding whatever teamwork-related aspect they deem important during the first five cases and also to re-evaluate the teamwork if needed. Instructors will stay with their groups during the whole simulation event to guarantee coordinated, consistent and longitudinal feedback.

Data collection

Upon arrival, every student participant will create an individual anonymised study code, which will be entered on every form and questionnaire and will allow us to link all measurements during the course of the night. Students will also track their role (leader, member, observer) after every case to allow subgroup analyses in relation to these roles. Figure 1 depicts the data collection procedure during the night-shift simulation.

Before starting the simulation, all 35 participants will be asked to fill in a first questionnaire that assesses possible confounders such as demographic data, professional training as a nurse or paramedic, or any training in teamwork/human factors. Next, students will be randomly assigned to seven groups via a computer-generated algorithm by the principal investigator. Four groups will serve as intervention groups and the remaining three

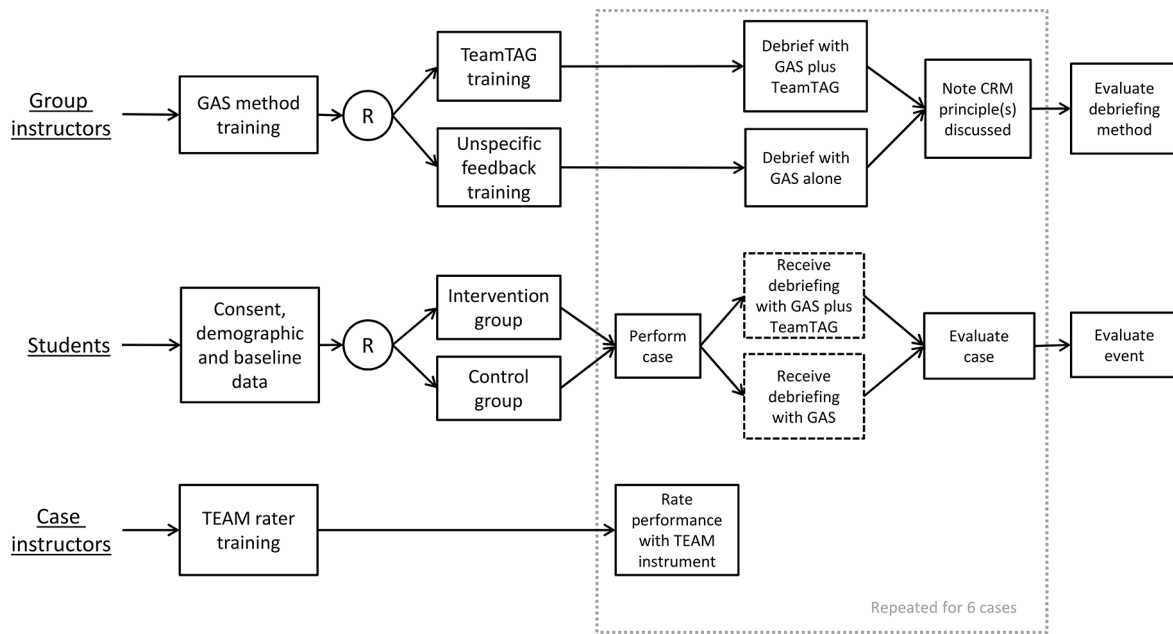


Figure 1 Study flow chart. CRM, crisis resource management; GAS, gather–analyse–summarise; R, randomisation; TEAM, Team Emergency Assessment Measure; TeamTAG, teamwork techniques analysis grid.

as controls; participants will not know to which condition they are assigned. After randomisation, all groups will gather separately and will be asked to discuss already known principles of teamwork and 15 multiple-choice questions concerning emergency medicine. A recent study showed that the results of such discussions are linked to team performance.⁴⁴

During the simulation, all groups will face six simulations where teamwork will be measured and teamwork-related feedback provided. All cases depict common emergency situations where the participation of an emergency team in the emergency room is needed. Table 1 gives a brief overview of the diagnoses of the six cases and challenges for teamwork.

During every case, team performance will be measured using the TEAM,¹⁶ which will be filled in by the case instructors and an additional rater. The two TEAM raters will be blind to the debriefing condition the group is assigned to.

After every case (duration about 30 min), debriefing will start (duration about 20 min) with checklist-based feedback from the simulated patients (focus: communication skills, empathy) and the case instructors and peer observers (focus: factual knowledge, diagnostic skills). As the last part of the debriefing process, the teamwork-related debriefing will be conducted by the group instructor using the GAS method with or without the support of the TeamTAG depending on the experimental condition. The strict timing, which will be centrally coordinated, will be necessary for a smooth transition of groups between cases and to ensure that the total length of the simulation does not exceed 8 hours.

After the debriefing process, all group members will be asked to evaluate the case and rate how helpful the debriefing was. Group instructors in both conditions will track the main topics of their teamwork debriefing in a debriefing protocol as free text. After the simulation, the content of these debriefing protocols will be clustered

Table 1 Teamwork-relevant cases presented in the emergency department simulation

Case	Diagnosis	Challenges for teamwork
1	Exacerbated COPD	Conflict management, control of emotions due to challenging patient
2	Ischaemic stroke of middle cerebral artery	Task management, communication with colleagues Manage aphasic patient
3	STEMI and non-sustained ventricular tachycardia	Patient deterioration (cardiac arrhythmia) during care
4	Ventricular fibrillation following STEMI	Team leadership, structured ACLS
5	Haemodynamically unstable ruptured spleen	Set priorities in evaluation and management, structured ATLS
6	Head laceration with ethanol intoxication	Manage agitated patient

ACLS, advanced cardiac life support; ATLS, advanced trauma life support; COPD, chronic obstructive pulmonary disease; STEMI, ST-elevation myocardial infarction.

independently (JF and FS) and matched with CRM principles.

Right after the last case of the night, all participants will fill in a final evaluation, which will ask them to list all the CRM principles on which they received feedback during the night. Participants will also evaluate the importance of each principle for their future work as physicians and provide a general evaluation of the night. Every group tutor will rate the feasibility, efficiency and difficulty of providing feedback.

Collected data

1. Baseline characteristics: The data collected on the first questionnaire and the results of group and teamwork discussions will be used to compare the baseline between the two conditions. Discussion results will be analysed qualitatively to identify differences in knowledge and in the personal definition of good teamwork at the beginning of the night. Furthermore, the TEAM scores during the first simulation case will serve as the baseline team performance.
2. Hypothesis 1 measurement (team performance, number of CRM principles discussed): Team performance will be evaluated using the 11 items of the translated TEAM. Similar to previous studies,^{15 16 41 45 46} we will analyse ratings on the item level (range: 0–4), the sum score (range: 0–44) and the overall rating per case (range: 1–10). The number of CRM principles discussed will be derived from two sources, namely, the debriefing protocols of the group instructors and participants.
3. Hypothesis 2 measurement (importance, satisfaction, helpfulness): Estimated relevance of the CRM principles learnt and overall satisfaction with the simulation will be evaluated on 7-point Likert scales at the end of the night. Helpfulness of the debriefing from the different providers (simulated patient, peer, case tutor and group tutor) will be rated by participants after every case on a 7-point Likert scale.
4. Hypothesis 3 measurement (instructor ratings): Debriefing evaluation of the group instructors (feasibility, efficiency and difficulty of providing feedback) will be measured with 7-point Likert scales and as free-text answers at the end of the night.
5. Other measures: The general evaluation form will ask participants to rate pleasure, quality of instruction during the night, difficulty of cases and possibility of applying knowledge on 7-point Likert scales.

All 7-point Likert scales will be coded from +3 (*strongly agree*) to –3 (*strongly disagree*). All data collection forms will be available upon request.

Analyses

Data will be analysed in SPSS 24 and R using descriptive, inferential and explorative statistics. We conducted a calculation of power for our primary research question (team performance). Recent studies, reporting mainly data for well-trained and experienced teams, showed

TEAM sum scores up to 40.^{45 46} Only one study provided data for less experienced teams with a TEAM sum score of 21.⁴⁵ On the basis of these results and data from a pre-study (see the TeamTAG section in the Preliminary results section), we expect a TEAM sum score of about 20 for an untrained team and a score of around 40 for teams that receive a training related to teamwork skills and/or have a lot of experience in this area. These scores indicate a potential increase due to training of up to 20 points on the TEAM sum score. As a relevant training effect for a single training event such as ours, we estimate a gain in the TEAM sum score of 11 points (ie, one point per item). Using the SD from the last published study on the TEAM⁴⁶ ($SD=4.4$) and $\alpha<0.05$, we have determined that about six teams are needed to detect a significant difference between the conditions with a power of 80%. Missing data will be handled using pairwise deletion.

1. Baseline characteristics: Discussion results of the intervention and control groups will be compared using qualitative methods and confounder analysis (demographics, prior training) with parametric and non-parametric tests for testing equivalence. The TEAM scores (single items, sum score, overall score) from the first simulation case will be compared between conditions using multilevel analyses to take the hierarchical structure of data into account.
2. Analyses for hypothesis 1: The TEAM scores (single items, sum score, overall score) of the intervention and control groups during the sixth simulation case will be compared using multilevel analyses. The development of team performance over the six cases will be analysed using descriptive statistics and plotting ‘training curves’ for each team. The total number of CRM principles discussed in the control and intervention groups will be compared using a multilevel model.
3. Analyses for hypothesis 2: The participants’ ratings of the feedback’s helpfulness, the importance of CRM principles and satisfaction with the debriefing will be compared between the control and intervention groups using multilevel models.
4. Analyses for hypothesis 3: Group instructors’ evaluations of the instrument will be examined descriptively.
5. Other measures: The general evaluation will be examined in a descriptive way.

Methodological limitations

Group instructors will not be observed while debriefing due to our limited labour force. Therefore, we cannot be sure the quality of the debriefing will be comparable among the seven participating groups. Further studies could use debriefing assessment tools such as the Observational Structured Assessment of Debriefing tool,⁴⁷ which might help distinguish between effects of overall debriefing quality and our approach. In our study, we will try to address this limitation with extensive group

instructor training to ensure an equal qualification level regarding debriefing and with a randomisation of instructors to conditions. Furthermore, participants will be asked to state the debriefing topic and to rate the quality of debriefing after every simulation case, which will be reported in later publications.

The time for debriefing after every case will be relatively short due to the design of our 8-hour simulation, where all groups will rotate through six cases to give participants a broad overview of emergency medicine and application areas of CRM. To use this limited time most productively, we have added additional specifications for debriefing (eg, focus on one or two principles per debriefing session, as described in the Methods and analysis section) because some instructors stated in a prestudy that the time allowed for debriefing was not sufficient. Future studies could investigate whether results of this study hold if all CRM principles are being discussed and thus repeated after every case/more often during the night and if time for debriefing is longer. Until now, there has been no strong evidence for the superiority of a longer debriefing.²¹

The study will focus only on short-term effects of two different debriefing approaches. Further research should investigate long-term effects on performance or changes in behaviour during clinical practice. A last limitation of this study is that it is a single-centre study and so results might be limited to local circumstances.

Data sharing statement

Data analysis will be conducted by the investigator's team (data management team). As the study is not a clinical trial, a data-monitoring team is not needed. The anonymised full data set will be published together with the journal publication or using the Dryad Data Repository (Durham, NC, USA) as required by the journal's guidelines. Data will furthermore be stored in the local data repository at Charité Medical School Berlin according to the local guidelines for good scientific practice.

PRELIMINARY RESULTS

Validation of the German TEAM

The German TEAM can be found in the online supplementary information. As a preliminary validation, inter-rater correlation was checked between three investigators (JF, FS and DE) and an external expert on two videotaped resuscitations. Both resuscitations were simulation based and had similar factual content; however, the first simulation showed good teamwork and the second intermediate teamwork performance. The videotaped simulations were used for group instructors' debriefing training and for validity testing of the German TEAM.

Intraclass correlation coefficients were .99 for the first resuscitation (mean TEAM score=42.3, $SD=1.3$) and .85 for the second (mean TEAM score=22.5, $SD=3.1$), which indicates excellent inter-rater agreement. For this reason, we consider the German TEAM a valid instrument for assessing team performance in our study.

TeamTAG

A first version of TeamTAG was used in a prestudy, conducted during the previous simulated night shift in 2016. In this prestudy, all instructors ($n=7$) used TeamTAG as part of their debriefing (similar to the GAS method plus TeamTAG). They were asked to rate the feasibility and helpfulness of the TeamTAG (7-point Likert scale; -3 to $+3$), as well as whether time for debriefing was sufficient (7-point Likert scale; -3 (*strongly insufficient*) to $+3$ (*strongly sufficient*)). Furthermore, they could comment on specific aspect of the guideline they liked or disliked (free-text answers). All participants were asked how useful the instructors' feedback was (7-point Likert scale; -3 to $+3$).

Instructors rated the guideline as a feasible tool ($M=1.9$, $SD=0.9$) and stated that it helped them in both observing and giving feedback to the participants of the simulation ($M_{\text{observe}}=2.3$, $SD=0.8$; $M_{\text{feedback}}=2.3$, $SD=0.5$). They had a heterogeneous view of the adequacy of time available for debriefing ($M=-0.3$, $SD=1.1$) The participants declared having found the feedback to be useful ($M=1.7$, $SD=1.0$).

ETHICS AND DISSEMINATION

The study protocol was designed according to the Declaration of Helsinki, the local guidelines for good scientific practice at Charité Medical School Berlin and the ICMJE (International Committee of Medical Journal Editors) recommendations. The study protocol was approved by the institutional office for data protection (AZ 737/16) and the ethics committee at Charité Medical School Berlin (EA2/172/16).

All participants and instructors will provide informed consent. Because the simulation is already a well-known event at Charité Medical School Berlin and receives official teaching funds, participants who refuse to take part in our study must have a chance to participate nevertheless. In this case, students will not provide the informed consent prior to randomisation; instead, an independent 'no-study' group will then be created, which will be identical to the control group but without any teamwork debriefing. We do not expect any harm for students who undergo the intervention.

Publication

Results of the study will be presented during national and international scientific meetings. The authors aim to publish all results in a peer-reviewed journal. Part of the protocol has been previously presented at the Research in Medical Education (RIME) conference in Dusseldorf, Germany, in March 2017 and was awarded the RIME Award: Best Research Protocol 2017.⁴⁸

Author affiliations

¹Simulated Patients Program, Charité Medical School Berlin, Berlin, Germany

²Lernzentrum (Skills Lab), Charité Medical School Berlin, Berlin, Germany

³Department of Emergency Medicine, Inselspital, University of Bern, Bern, Switzerland

⁴Department of Anesthesiology and Operative Intensive Care Medicine CCM & CVK, Charité Medical School Berlin, Berlin, Germany

⁵Progress Test Medizin, Charité Medical School Berlin, Berlin, Germany

⁶Max Planck Institute for Human Development, Center for Adaptive Rationality, Berlin, Germany

Acknowledgements The authors would like to acknowledge Hanno Heuzeroth and David Steinbart for support in conducting the pre-study and Simon Cooper for support in using and translating the TEAM. Furthermore, the authors thank all instructors and physicians, especially Tobias Deselaers, for organising the simulation and their willingness to participate as raters and instructors, and such. In addition, we would like to acknowledge Jane Runnacles (London), Suzanne Bentley (New York) and Christopher Timmis (Wolverhampton) for their constructive critique on an earlier version of the manuscript. The authors thank Anita Todd for editing the manuscript.

Contributors JF and FS translated the TEAM, designed the study and will be responsible for conduction. DE, WEH and JEK contributed to the study design. JEK supervised the study design and will supervise conduction. JF and FS are responsible for data analyses. JF, FS and JEK wrote the manuscript. JF and FS conducted the pre-study. DE is responsible for funding and local administration at Charité Medical School Berlin and heads the steering committee. All authors carefully read the manuscript, made critical and substantial revisions, and gave their approval for publication.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. The investigators JF and FS are funded by the German Federal Ministry of Research and Education (BMBF). The sponsor did not interfere with the conception and conduction of the study, data analysis or production of the manuscript.

Competing interests WEH received financial compensation for educational consultancy from the AO Foundation, Zurich, Switzerland. All other authors report no competing interests.

Ethics approval Ethics committee of Charité Medical School Berlin.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- de Vries EN, Ramrattan MA, Smorenburg SM, *et al.* The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care* 2008;17:216–23.
- Kohn LT, Corrigan JM, Donaldson MS, eds. *To err is human: building a Safer Health System*. Washington: National Academies Press, 2000.
- Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008;121(5 Suppl):S2–S23.
- Schwartz A, Elstein AS. *et al*/Clinical reasoning in medicine. In: Higgs J, Jones M, Loftus S, Christensen N, . eds. *Clinical reasoning in the health professions*. 3rd ed. Oxford (UK), 2008:223–34.
- Hautz SC, Schuler L, Kämmer JE, *et al.* Factors predicting a change in diagnosis in patients hospitalised through the emergency room: a prospective observational study. *BMJ Open* 2016;6:e011585.
- Risser DT, Rice MM, Salisbury ML, *et al.* The Potential for Improved Teamwork to Reduce Medical Errors in the Emergency Department. *Ann Emerg Med* 1999;34:373–83.
- Leape LL. Error in medicine. *JAMA* 1994;272:1851–7.
- Hautz WE, Kämmer JE, Exadaktylos A, *et al.* How thinking about groups is different from groupthink. *Med Educ* 2017;51:229.
- Hautz WE, Kämmer JE, Schaubert SK, *et al.* Diagnostic performance by medical students working individually or in teams. *JAMA* 2015;313:303–4.
- Knaus WA, Draper EA, Wagner DP, *et al.* An evaluation of outcome from intensive care in major medical centers. *Ann Intern Med* 1986;104:410–8.
- Manser T. Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. *Acta Anaesthesiol Scand* 2009;53:143–51.
- Frank J, Snell L, Sherbino J, eds. *CanMEDS 2015 Physician Competency Framework*. Ottawa: Royal College of Physicians and Surgeons of Canada, 2015.
- Kaba A, Wishart I, Fraser K, *et al.* Are we at risk of groupthink in our approach to teamwork interventions in health care? *Med Educ* 2016;50:400–8.
- Kiesewetter J, Gutmann J, Drossard S, *et al.* The Learning Objective Catalogue for Patient Safety in Undergraduate Medical Education—A Position Statement of the Committee for Patient Safety and Error Management of the German Association for Medical Education. *GMS J Med Educ* 2016;33:1–14.
- Cooper S, Cant R, Connell C, *et al.* Measuring teamwork performance: Validity testing of the Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams. *Resuscitation* 2016;101:97–101.
- Cooper S, Cant R, Porter J, *et al.* Rating medical emergency teamwork performance: development of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2010;81:446–52.
- Ochsmann EB, Zier U, Drexler H, *et al.* Well prepared for work? Junior doctors' self-assessment after medical education. *BMC Med Educ* 2011;11:99.
- Hayes CW, Rhee A, Detsky ME, *et al.* Residents feel unprepared and unsupervised as leaders of cardiac arrest teams in teaching hospitals: a survey of internal medicine residents. *Crit Care Med* 2007;35:1668–72.
- Brydges R, Hatala R, Zendejas B, *et al.* Linking simulation-based educational assessments and patient-related outcomes: a systematic review and meta-analysis. *Acad Med* 2015;90:246–56.
- Morgan L, Hadi M, Pickering S, *et al.* The effect of teamwork training on team performance and clinical outcome in elective orthopaedic surgery: a controlled interrupted time series study. *BMJ Open* 2015;5:e006216.
- Cheng A, Eppich W, Grant V, *et al.* Debriefing for technology-enhanced simulation: a systematic review and meta-analysis. *Med Educ* 2014;48:657–66.
- Gaba D, Fish K, Howard S. *Crisis Management in Anesthesiology*. New York: Churchill Livingstone, 1994.
- Morey JC, Simon R, Jay GD, *et al.* Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MedTeams project. *Health Serv Res* 2002;37:1553–81.
- DeVita MA, Schaefer J, Lutz J, *et al.* Improving medical emergency team (MET) performance using a novel curriculum and a computerized human patient simulator. *Qual Saf Health Care* 2005;14:326–31.
- Theilen U, Fraser L, Jones P, *et al.* Regular in-situ simulation training of paediatric Medical Emergency Team leads to sustained improvements in hospital response to deteriorating patients, improved outcomes in intensive care and financial savings. *Resuscitation* 2017;115:61–7.
- Sawyer T, Eppich W, Brett-Fleegler M, *et al.* More Than One Way to Debrief: A Critical Review of Healthcare Simulation Debriefing Methods. *Simul Healthc* 2016;11:209–17.
- van de Ridder JM, Stokking KM, McGaghie WC, *et al.* What is feedback in clinical education? *Med Educ* 2008;42:189–97.
- Cook DA, Hamstra SJ, Brydges R, *et al.* Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Med Teach* 2013;35:e867–e898.
- Phrampus PE, O'Donnell JM. *et al*/Debriefing Using a Structured and Supported Approach. In: Levine AL, DeMaria SJ, Schwartz AD, . eds. *The Comprehensive Textbook of Healthcare Simulation*. New York: Springer, 2013:p.73–84.
- Cheng A, Rodgers DL, van der Jagt ?lise, *et al.* Evolution of the Pediatric Advanced Life Support course. *Pediatric Critical Care Medicine* 2012;13:589–95.
- Stanford School of Medicine. What Are Cognitive Aids? Cognitive Aids in Medicine . 2016 <http://cog aids.stanford.edu/whatarecog aids.html> (cited 2016 Dec 14).
- Harrison TK, Manser T, Howard SK, *et al.* Use of cognitive aids in a simulated anesthetic crisis. *Anesth Analg* 2006;103:551–6.
- Marshall S. The use of cognitive aids during emergencies in anesthesia: a review of the literature. *Anesth Analg* 2013;117:1162–71.
- Haugen AS, Sjøfeland E, Almeland SK, *et al.* Effect of the World Health Organization checklist on patient outcomes: a stepped wedge cluster randomized controlled trial. *Ann Surg* 2015;261:821–8.
- Cheng A, Hunt EA, Donoghue A, *et al.* Examining pediatric resuscitation education using simulation and scripted debriefing: a multicenter randomized trial. *JAMA Pediatr* 2013;167:528–36.

36. Stegmann K, Pilz F, Siebeck M, *et al.* Vicarious learning during simulations: is it more effective than hands-on training? *Med Educ* 2012;46:1001–8.
37. Weinman J. *Psychology as applied to medicine*. 2nd ed. Bristol, England: Wright, 1987.
38. Jenkins V, Fallowfield L. Can communication skills training alter physicians' beliefs and behavior in clinics? *J Clin Oncol* 2002;20:765–9.
39. Stroben F, Schröder T, Dannenberg KA, *et al.* A simulated night shift in the emergency room increases students' self-efficacy independent of role taking over during simulation. *BMC Med Educ* 2016;16:177.
40. Rall M, Oberfrank S. 'Human factors' und 'crisis resource management'—Erhöhung der Patientensicherheit. *Unfallchirurg* 2013;116:892–9.
41. Couto TB, Kerrey BT, Taylor RG, *et al.* Teamwork Skills in Actual, In Situ, and In-Center Pediatric Emergencies. *Simulation in Healthcare* 2015;10:76–84.
42. Dorer B. ESS Round 7 translation guidelines [Internet]. *European Social Survey* 2014. http://www.europeansocialsurvey.org/docs/methodology/ESS_R7_Translation_Guidelines_FINAL.pdf (cited 2017 Jan 5).
43. Feldman M, Lazzara EH, Vanderbilt AA, *et al.* Rater training to support high-stakes simulation-based assessments. *J Contin Educ Health Prof* 2012;32:279–86.
44. Park HR, Kim CJ, Park JW, *et al.* Effects of team-based learning on perceived teamwork and academic performance in a health assessment subject. *Collegian* 2015;22:299–305.
45. Maignan M, Koch FX, Chaix J, *et al.* Team Emergency Assessment Measure (TEAM) for the assessment of non-technical skills during resuscitation: Validation of the French version. *Resuscitation* 2016;101:115–20.
46. Cant RP, Porter JE, Cooper SJ, *et al.* Improving the non-technical skills of hospital medical emergency teams: The Team Emergency Assessment Measure (TEAM™). *Emerg Med Australas* 2016;28:641–6.
47. Arora S, Ahmed M, Paige J, *et al.* Objective structured assessment of debriefing: bringing science to the art of debriefing in surgery. *Ann Surg* 2012;256:982–8.
48. Freytag J, Stroben F, Kämmer JE, *et al.* Teamwork-related debriefings in undergraduate emergency-medicine simulations : evaluation of the 'TeamTAG' using a prospective- randomized pilot-study design. *Paper Presented at 5th International Conference for Research in Medical Education (RIME 2017)*. Düsseldorf:German Medical Science GMS Publishing House, 2017