Generic Crew Resource Management training to improve non-technical skills in acute care - phase 2 : a pre-post multicentric intervention study

Generic Crew Resource Management training to improve non-technical skills in acute care - Phase 2: A pre-post multicentric intervention study

ABSTRACT

BACKGROUND Crew Resource Management (CRM) training aims to improve non-technical skills to reduce preventable errors in healthcare. This study evaluates a generic CRM training program for acute care settings by assessing its' effect on the four levels of Kirkpatrick.

METHODS The intervention entails a theoretical part and a simulation-based multidisciplinary team training including debriefing. Pre-post measurements were taken using questionnaires, observations and interviews. 231 Belgian physicians, midwives and nurses participated.

RESULTS On level 1, respondents provided high scores and perceived the training as important. On level 2, a shift in knowledge, skills and awareness was found for 9/23 items. Observations showed safer teamwork on level 3 for all specialty areas. On level 4, a significant improvement was measured only for perceptions of management. Hierarchy and lack of a safety culture hinder the application of CRM in practice.

CONCLUSIONS A generic CRM-training invokes a positive reaction, changes attitude, knowledge, non-technical skills, and behavior. A onetime CRM-training is insufficient to change safety culture. The possibility of a generic training for acute specialty areas was confirmed.

KEY POINTS

- A generic CRM-training does not only invoke a positive reaction and changes attitude, knowledge and non-technical skills, but also objectively changes behavior
- A single CRM-training of the entire unit is insufficient to change safety culture
- A generic training for acute specialty areas (OB, ER, ICU) is possible

KEYWORDS Crew Resource Management, CRM, Non-technical skills, Team training, Interdisciplinary, Simulation, Human Factors

BODY TEXT

INTRODUCTION

Since the seminal publication of the Institute of Medicine and the WHO Curriculum for Patient Safety, it is increasingly acknowledged that non-technical skills or human factors are an important source of errors in healthcare affecting many patients worldwide (Institute of Medicine, 2000; Franck, Roes, De Schepper, & Timmermans, 2018; Herzberg et al., 2019; James, 2013). Research demonstrates that a substantial amount of these errors are preventable and deficits in non-technical skills, such as sub-optimal teamwork, interprofessional communication and decision making, or unclear leadership and task coordination, remain key root causes (Fransen et al., 2012; Guise & Segel, 2008; Higham & Baxendale, 2017; Hull et al., 2012; Kao & Thomas, 2008; Leonard, Graham, & Bonacum, 2004; Lingard et al., 2004; Rall & Dieckmann, 2005).

Improving healthcare safety prompts comparison with other high risk industries. In aviation, research identified that over 70% of aviation accidents are due to human factors. Consequently, teamwork skill training programs called Crew Resource Management training have become mandatory for flight crews worldwide. They are focused on training non-technical skills, or the cognitive, social and self-management skills that contribute to safe and efficient task performance (Flin, O'Connor, & Crichton, 2008; Willems, Kurka, Bohmann, Rostek, & Pfeilschifter, 2019).

Because of the similarities Crew Resource Management (CRM) training formats were adapted to healthcare environments (Gross et al., 2019) with the aim of reducing potentially preventable errors (Chen, Iqbal, & Li, 2017; Joint Commission International, 2014; Kohn, Corrigan, & Donaldson, 2000; Manser, 2009; Murphy, 2006; Sundar et al., 2007). Although the first efforts in developing a CRM training for healthcare professionals date from the early 1980’s in the field of anesthesia, concerns about the delivery and evaluation of CRM programs since then can be raised. A recent review of Gross and colleagues (2019) on CRM training in healthcare highlighted the need for more research in

order to establish non-educational criteria for success in the implementation of CRM in healthcare organizations (Gross et al., 2019). The present study wants to contribute to the literature, by evaluating a generic Crew Resource Management training approach for healthcare professionals working in acute care settings. By this we mean an equal process of training between the settings. However, the cases used in the training were department-specific. Such a well defined generic CRM team training program could facilitate the implementation of CRM training across acute care settings.

THEORETICAL FRAMEWORK

First, the development of such a generic CRM training must be guided by a team task analysis. Therefore, a preceding needs assessment survey mapped the knowledge, skills, attitudes and concerns of various acute care professionals toward non-technical skills and a generic CRM training. The results indicate that a generic CRM training program for all acute care professionals is feasible provided that specific barriers are taken into account. Creating a safe no-blame learning environment proved key and a theoretical part preceding the simulation-based training must help raise awareness of human factors and patient safety (*reference removed for blinding*).

Second, as a generic CRM training program for acute care is a new approach, it needs to be extensively evaluated. Although various models exist, the best-known and most widely used model for training evaluation is the four-level approach of Kirkpatrick (Kirkpatrick, 1979). The strength of this model lies in its simplicity and pragmatic way of helping practitioners think about training programs (Tamkin, Yarnall, & Kerrin, 2002). Although this model dates back to 1959, Tamkin, et al. (2002) state in their review that whilst a diversity of terminology and categories exist, there are huge areas of similarity in the range of evaluation models on offer. In addition, evaluation strategies do not appear to have changed significantly since the development of Kirkpatrick's model (Tamkin et al., 2002). Kirkpatrick's model consists of four stages or levels which have no causal or sequential relationship (Kirkpatrick & Kirkpatrick, 2016):

1) Level 1: Reaction – what the participants think of the program

2) Level 2: Learning – the changes in knowledge, skills, or attitude with respect to the training objectives
3) Level 3: Behavior – changes in job behavior resulting from the program, to identify whether the learning is being applied
4) Level 4: Results – the bottom-line contribution of the training program

The aim of the study is to evaluate a generic CRM training program for acute care settings by assessing its' effect on the four levels of Kirkpatrick. Additionally, the generic nature of the training program is evaluated by comparing these effects between several acute care specialty areas.

METHODS

METHODOLOGY

An intervention study with a pre- post measurement was set up.
Two research questions were formulated:

1) What is the effect of a generic CRM training program in acute care settings on the different levels of Kirkpatrick?
    a. Participant reaction to the training
    b. Participant attitude, knowledge and non-technical skills
    c. Participant behavior in clinical practice
    d. Changes on organizational level (Kirkpatrick & Kirkpatrick, 2005)
2) Is there a difference in effect of the generic CRM training program between acute care specialty areas (emergency rooms, intensive care units and obstetric units)?

SETTING AND PARTICIPANTS

The study and training conditions (multidisciplinary team training in situ, training period, data collection techniques) were communicated to all acute care hospitals in the Dutch

speaking part of Belgium. Fourteen acute care wards were prepared to participate in the study between April 2016 and June 2017. This included eight emergency rooms, one intensive care unit and five obstetric units. In total 508 physicians, midwives and nurses participated in the study. Anesthesiologists and all care professionals other than midwifes/nurses and physicians were excluded.

DESCRIPTION OF THE INTERVENTION

Based on the results of the preceding needs assessment survey *(reference removed for blinding)* and a literature study, a generic CRM training was developed. This training program consisted of two parts: a theoretical part and a simulation-based multidisciplinary team training including debriefing. The theoretical part aimed to raise awareness on human factors and patient safety and provide the participants with knowledge concerning non-technical skills and how to apply these during emergency situations. It was offered as a plenary session (1.5h) or as an e-learning program (0.5h).

For the simulation-based team training, acute care professionals were divided into groups of four to six. At least one physician and three midwives/nurses had to be available for the training to take place. The three-hour simulation-based training was organized by two EUSIM-trained facilitator/researchers (European simulator instructor course). The training was provided on the participants' own ward to ensure familiarity with materials and stimulate realistic behavior.

Each training started with a half-hour orientation moment to explain the purpose of the training, the importance of a safe learning environment and the familiarization with the patient simulator and materials (i.e. telephone system). Then, a genuine emergency situation was simulated (i.e. eclampsia on an obstetric ward). Participants were briefed on the starting point of the simulated scenario and the current condition of the patient. In the emergency room and intensive care unit we used a mannequin-based simulation, in the obstetric unit we used a real-life simulant. In this high-fidelity simulations, the working environment was mimicked and served as an educational risk-free training tool (Carron,

Trueb, & Yersin, 2011). The team applied their usual and known protocols to deal with the simulated emergency situation. The simulation itself took about 20 minutes and always ended in the team gaining control over the situation. Each simulation was followed by a debriefing of on average 45 minutes. During this debriefing participants reflected on their actions and non-technical skills in the simulation. To protect the safe learning environment facilitators emphasized positive feedback and mutual points of improvement concerning collaboration and communication. Subsequently, a second emergency situation (i.e. neonatal resuscitation on an obstetric ward) was simulated and debriefed. The need for this second scenario derived from revision of the CRM training by a group of experts and prototyping by an entire multidisciplinary emergency room team. As such, participants had the opportunity to apply and internalize what they learned from the first scenario.

This simulation-based training in small groups was repeated until every physician, midwife/nurse on the acute care ward had been trained.

DATA COLLECTION

For each of the levels of Kirkpatrick a data collection technique was matched to ensure valuable data. This resulted in a mix of three techniques: online questionnaires (Qualtrics®), observations and interviews. Pre- and post measurements were obtained one month preceding the program and one month after the program for the questionnaires and observations of levels two to four.

The observations were executed during unexpected simulations on the ward. The time of measurement and the scenarios were kept identical. Each ward had one unexpected simulation at pretest and one unexpected simulation at posttest. Therefore, the pre- and posttest teams did not necessarily consist of the same individuals as only the team members working at that specific time of measurement were observed. We considered this random selection of team members to be representative of their team. In total, eighty-six participants were observed. Although these observations were executed in all 14 acute

care wards, the film of two wards was damaged. Therefore, only 12 pre-post observations could be analyzed. An expert rater judged the non-technical team skills in every film in random order without knowing if the film was a pre- or posttest.

Following the training permission was asked to contact participants for an interview. From this pool a purposive sample was selected based on participation in both theory and simulation, and a variation in age and function. Eleven OB and ER participants were interviewed within one month after the training. Each semi-structured interview began with the open and broad question: "What did you think of the training topic?". During the interview the researcher posed probing questions to uncover the participant's experience with the training on each of the levels of Kirkpatrick. The interviews were audiotaped.

An overview of the data collection techniques and instruments can be found in Table 1.

TABLE 1 NEAR HERE

STATISTICAL ANALYSIS

Statistical data analyses were conducted using SPSS version 26.0 (IBM, 2019). Only respondents who followed the entire program (theory and simulation) and completed both the pre- and posttest were included in the analysis. This resulted in a sample of 231 respondents.

Quantitative analysis

Level 1 – participant reaction to the training

Descriptive statistics were provided for the program evaluation scores and EOC questions that were only posed at posttest. The McNemar testing was carried out to detect significant changes between EOC questions that were posed pre- and posttest.

Level 2 – participant knowledge and non-technical skills

The HFAS questions were originally set on a 5-point Likert scale, which was recoded to construct a binary variable: totally disagree or disagree = 0 ; totally agree and agree = 1, and no opinion was set to missing. McNemar testing was carried out to detect significant changes between pre- and posttest measurements.

Level 3 – participant behavior in clinical practice

The pre- and posttest was considered a paired measurement, which was analyzed in a linear mixed model framework. Across all models, the score on the Clinical Teamwork Scale was entered as dependent variable, time (pre or post) as fixed effect, and setting as random intercept. For characteristic comparison of the pre- and post intervention teams participating in the unexpected simulations Chi-square or Kruskal-Wallis tests were performed (Table 3).

Additionally, we tested a main effect of specialty area (emergency rooms, intensive care units and obstetric units) in a model with time and specialty area as fixed effects.

Level 4 – changes on organizational level

The SAQ questions were originally set on a 5-point Likert scale, which was recoded to construct a binary variable: totally disagree or disagree = 0 ; totally agree and agree = 1, and no opinion was set to missing. The summary values of the SAQ were treated as numerical (continuous) outcomes. The difference between the pre- and the posttest was modeled by fitting mixed models with individual ID as random effect, to account for the repeated measurements within the same individual. Time was entered as a fixed effect. Due to convergence problems, it was not possible to include the specialty area as covariate in the analysis.

Across all analyses, a P-value below0.05 was considered significant.

Qualitative analysis

The interviews were transcribed verbatim. A descriptive thematic analysis per level of Kirkpatrick was performed with themes emerging from the data. The software program

NVIVO 12 was used. For dependability focusing on the research objective, trying to explore the same areas for all the participants and self-reflection were important points of interest.

ETHICAL CONSIDERATIONS

Data were collected and analyzed confidentially, taking into account European legislation regarding the "General Data Protection Regulation" (GDPR). Because this concerns a study in which only adult healthcare workers participate on their own free will and after informed consent, based on the ICH-GCP principles ethical approval was not sought for the present study (European Medicines Agency, 2016). Furthermore, the management of every participating hospital approved the study and every potential respondent received an invitational letter containing information on the study objective and methodology, and informed consent was obtained at the beginning of the study. Additionally, participants were allowed to terminate the study at any time they desired.

RESULTS

A description of the respondents is provided in Table 2.

TABLE 2 NEAR HERE

LEVEL 1 – PARTICIPANT REACTION TO THE TRAINING

Quantitative results - General

Figure 1 depicts high overall scores on the separate training parts. In addition, participants provided high usefulness scores for both the theoretical and the simulation-based training part. This indicates that what they learned was strongly transferable to their daily practice. The mean score of the entire training program (theoretical and simulation-based part) totaled up to 7.4/10 (SD 1.6).

FIGURE 1 NEAR HERE

Following the training program the majority of participants was convinced that this training could improve patient safety and quality of care (80%). It caused them to act differently in acute situations (79%) and improved their knowledge and skills (68%). More than half of participants stated that they really needed this training (57%). Table 3 depicts how their attitude towards simulation-based training changed after the intervention.

TABLE 3 NEAR HERE

Quantitative results – comparison between acute care specialty areas

Although high scores were noted for all specialty areas, ER participants scored the theoretical part, the simulation-based part and the entire training program significantly lower than participants of obstetric or intensive care units. As such, the mean score of the entire training program totaled up to 7.0/10 (SD 1.8), 7.8/10 (SD 1.3), and 8.3/10 (SD 0.9) for ER, OB and ICU respectively ($p < 0.001$).

In line with these results, at post test ER participants indicated significantly less that the training caused them to act differently in acute situations (73% versus 88% in OB and 82% in ICU, $p=0.025$), improved their knowledge and skills (61% versus 78% in OB and 82% in ICU, $p=0.014$), and that they really needed this training (50% versus 65% in OB and 82% in ICU, $p=0.019$).

However, when comparing participant's attitude toward simulation-based training before and after the intervention, some of the greatest significant improvements after training are observed in emergency rooms. For instance, ER participant's worry about how others will judge their work based on how they perform during simulation-based training

decreased 20% at post test ($p<0.001$), while OB noted a reduction of 1% ($p= 1.000$) and ICU a steady state ($p=1.000$).

Qualitative results

The overall perception of participants was that the training was important, well designed and left them with a positive feeling. The scenarios in the simulation-based part were perceived as not too technical, making it possible to focus on the communication.

Some found the scenarios realistic, others stated that they did not reflect a real life situation. For the latter, the mannequin often proved to be a stumbling block.

> "The scenarios were very realistic. Especially the eclamptic convulsions." (Interviewee 3 – OB, Midwife)

> "This is not a real life situation. From a real patient you get feedback. It is very difficult for people who have seen this in real life. In a real situation I would have done different things much faster." (Interviewee 6 – ER, physician)

Some participants were anxious to participate in the training because they were afraid to fail. Others were looking forward to training with their team because they felt safe. Overall they described that they had to cross an initial threshold of anxiety and resistance, but had a changed attitude after participating.

> "Some of the older generation asked "do we really have to do this?" But once we had done it, everyone was positive. We had the feeling that we had to maintain this and could do it more often." (Interviewee 7 – OB, Midwife)

LEVEL 2 – PARTICIPANT KNOWLEDGE AND NON-TECHNICAL SKILLS

Quantitative results - General

Participant's responses revealed a significant shift in 9 out of 23 items on knowledge, skills and awareness concerning human factors and non-technical team skills (Table 4).

TABLE 4 NEAR HERE

Quantitative results – comparison between acute care specialty areas

In 17 out of 23 items all three specialty areas displayed a similar increasing or decreasing trend.

For instance, 92% of OB participants agreed at posttest that team members should question the decisions or actions of the team leader during a procedure (item 14). This is an increase of 15% compared to the pretest (p=0.002). 88% of ER participants agreed, which indicates an increase of 12% (p = 0.008) and 91% of ICU participants, showing an increase of 9% (p=1.000).

In seven of these 17 items one specialty area displayed a steady state. For instance, 96% of OB participants agreed at posttest that the team leader should encourage team members to raise questions during emergencies (item 11). This is an increase of 4% compared to the pretest (p=219). 96% of ER participants agreed, which indicates an increase of 7% (p = 0.064) and 100% of ICU participants, displaying a steady state (p=1.000).

In six out of 23 items one specialty area displayed a contrasting trend. For instance, 92% of OB participants agreed at posttest that the team formation and decision-making skills of the team leader are as important as the technical skills (item 6). This is a decrease of 2% compared to the pretest (p=0.754). 92% of ER participants agreed, which indicates an increase of 6% (p=0.152) and 91% of ICU participants, showing an increase of 9% (p=1.000).

However, for both the steady state as the contrasting trends, the percentage of agreement at posttest remained within the same order of magnitude across the three specialty areas.

Qualitative results

Participants indicated that the training created more connection and understanding within the team. Additionally, it made them more aware of the importance of communication and teamwork.

> "The training was an eye-opener. I think that beforehand a lot of people thought: 'Do we need this? It is going fine.' But after the training you think: 'Yes, we do need this. It is not going fine.'" (Interviewee 1 – ER, Physician)

It also revealed points of improvement.

> "I look differently at teamwork now. Especially the factor of hierarchy that you have to overcome and dare to breach." (Interviewee 4 – OB, Midwife)

Participant's knowledge of CRM principles was focused on repeating and confirming instructions, time-out, speak-up and flexible leadership.

> "I have learned to speak up. Really talk more and think and act less to yourself. // communicate more openly. Do not assume that others think what you think or know what you think." (Interviewee 11 – OB, Midwife)

> "The most important thing that I remember is that leadership should be dealt with flexibly. Also in acute situations in healthcare." (Interviewee 9, ER, Nurse)

LEVEL 3 – PARTICIPANT BEHAVIOR IN CLINICAL PRACTICE

Quantitative results - General

Observations showed that participants exhibited different (and safer) teamwork during the unexpected simulation one month after the intervention compared to the behavior one month before the intervention. This change on the Clinical Teamwork Scale proved to be significant ($p = 0.003$; Wilcoxon Signed Rank test). On average, the post-intervention score was 1.08 units higher compared to the pre-intervention score (95%CI 0.63 - 1.53).

FIGURE 2 NEAR HERE

To exclude that the difference between the pre- and the post-intervention score was attributable to a difference in the composition of the intervention teams, we tested the pre- and the post teams for differences in several variables. None of these variables showed a difference between pre and post-intervention teams. It is therefore very unlikely that the difference in score between the pre- and post-intervention, is attributable to a difference in team composition.

TABLE 5 NEAR HERE

Quantitative results – comparison between acute care specialty areas

We tested a main effect of specialty area on the Clinical Teamwork Score in a model with time and specialty area (OB, ER, ICU) as fixed effects. The main effect of the specialty areas was not significant (p=0.380). Therefore, there seems to be no systematic difference in Clinical Teamwork score (across time points) between the different acute care specialty areas.

Qualitative results

Participants described that they noticed small changes in team behavior such as giving clear instructions and repeating them, short discussions after an intervention and speaking their mind. However, others stated that they did not notice any changes or had not yet had an opportunity to apply their knowledge.

LEVEL 4 – CHANGES ON ORGANIZATIONAL LEVEL

Quantitative results – General

Concerning the culture within the ward and organization, a significant improvement was measured only for the subcategory of healthcare professionals perceptions of management (p = 0.004).

TABLE 6 NEAR HERE

Quantitative results – comparison between acute care specialty areas

Due to convergence problems, it was not possible to include the specialty area as covariate in the analysis.

Qualitative results

Participants appreciated that their nurse manager was open to the training and made it mandatory for every team member.

> "The training was mandatory for our team. You could not escape it and actually I think that is very good." (Interviewee 1 – ER, Physician)

On organizational level participants indicated that hierarchy forms an important barrier to flexible leadership, speak-up and time-out.

> "You are in some kind of rank. And that is the largest stumbling block. Those people (physicians) are trained for it, but if you are 100% convinced of something, you should be able to open you mouth." (Interviewee 4 – OB, Midwife)

Additionally, lack of a safety culture hinders the application of CRM in practice.

> "For the smallest mistake you get slapped on the fingers and then stomped on 10 more times. That is how it is. That is the problem here. They are trying to change, but that's not easy." (Interviewee 6 – ER, Physician)

To overcome these obstacles, participants suggest a more thorough implementation of CRM, a larger framework, embedding it in existing technical trainings and structures, repeating the CRM training on a regular basis, broadening it to other wards, disciplines or the entire hospital and implementing debriefing after real life emergencies.

"I think it is situated within a larger problem. There should be a shift in mentality and everyone should have a simulation training once a year." (Interviewee - ER, Physician)

DISCUSSION

This study shows that an interdisciplinary CRM training is an effective strategy to improve teamwork. This was confirmed with significant improvements and qualitative data on three of the four levels of Kirkpatrick.

It is an unique study because of the generic character and 'in situ' format of the training, and the fact that it covers all four levels of Kirkpatrick.

Generic character

This study confirms the possibility of a generic training for acute specialty areas. Between the three specialty areas only limited differences were detected. Nevertheless, some nuances can be made. For instance on level 1 ER participants indicated a significantly lower scores for the theoretical part, the simulation-based part, the entire training program and their attitude towards the training. However, when comparing the pre- and posttest scores, ER also showed the largest changes in attitude. Therefore, it is possible that ER participants just had more room for positive changes.

In situ training format

The study is innovative as the CRM-training was provided in the team's ward (in situ simulation) as recommended in the literature (Fransen et al., 2017; Siassakos, Crofts, Winter, Weiner, & Draycott, 2009; Siassakos et al., 2013). Therefore, participants were able to use familiar materials and actually imagine the situation (Crofts et al., 2008). In the obstetric units patient-actors played the role of the patient and her partner to make the situation more lifelike. After all, in obstetric emergencies the patient is conscience and her partner and/or family present. Therefore, good communication with the patient and her entourage is crucial (Siassakos et al., 2009). In ICU and ER low-fidelity mannequins were used as the patients in the scenarios were unconscious and patient-actors were less desirable due to the limitations of treatment (resuscitation, …). However, several ER participants indicated in the interviews that the scenarios did not reflect a real life situation. For them the mannequin often proved to be a stumbling block.

Therefore, it might be rewarding to overcome the practical objections of transporting and installing high-fidelity mannequins on these wards. A mobile simulation lab with all material in flight cases could make this feasible.

All four levels of Kirkpatrick

A couple of studies have evaluated CRM-training on the higher levels of Kirkpatrick. However, most studies are limited to specific unit types and evaluation of participant's reactions to the training (Level 1) and the learning effect (Level 2) (Haller et al., 2008; O'Dea, O'Connor, & Keogh, 2014). A meta-analysis of 20 CRM intervention studies showed positive reactions and the learning effect of participants (O'Dea et al., 2014).

Exceptionally, Guise et al. (2008) evaluated the behavior of participants as well using observation. In line with our findings, they described positive results with teamwork scores increasing from 6.0 to 7.5 ($p = 0.014$) on the Clinical Teamwork Scale.

However, the study at hand could only show a small change (perception of management) in patient safety culture as measured with the SAQ questionnaire. Several arguments can be made. First of all, the questionnaire was completed shortly after the training. It can be assumed that changing a safety culture requires some time, which makes it unlikely to measure changes one month after training. On the other hand, it avoided time related effects that could occur when the measurement is postponed (O'Dea et al., 2014). Therefore, literature indicates that it is desirable to evaluate on multiple occasions (Fransen et al., 2017). Due to time constraints this was not possible.

Second, it could be questioned if the SAQ is the right questionnaire to evaluate level 4 of the CRM-training. After all, several variables of the questionnaire such as working conditions and job satisfaction were not included in the CRM-training. Therefore, it is very unlikely that the training would have influenced these variables. Several other studies also applied the SAQ for evaluating CRM training and found the same weak or absent effects (Haller et al., 2008). As such, it is important for future research to consider

parameters that align better with the content of CRM-training (O'Dea et al., 2014). Patient outcomes could be used. However, due to the heterogeneity of the patient populations in this study, it was not achievable. In addition, patient outcomes should be treated with caution as variables such as length of stay, are also too distant from the content of CRM-training and influenced by many other factors, making it impossible to measure an impact. Better outcome parameters would be the use of protocols or debriefings (O'Dea et al., 2014).

A third possibility for the limited effect on level 4 and also a restriction of the study is that the intervention was limited to a single training of the entire unit. Participants indicated this clearly in the interviews and suggested a more thorough implementation of CRM, including repeating the training on a regular basis. A yearly training was suggested. Future research should determine the optimal frequency of the training to guarantee a lasting effect (Fransen et al., 2017).

Limitations

The study uses a mix of data collection techniques to evaluate each level of Kirkpatrick providing various insights. However, a sequential explanatory design would have provided more depth to the findings as the quantitative data could be explained by the qualitative ones.

Furthermore, the study compares between three acute care specialty areas, yet it included only one ICU resulting in 20 trained ICU participants, 11 completed pre- and post-questionnaires and no interview participants. This was due to the complexity and work load of the study, resulting in less ICUs being prepared to participate. Despite the small sample size, similarities across all three acute care specialty areas could be found.

CONCLUSIONS

A generic CRM-training, as delivered in the present study, does not only invoke a positive reaction and changes attitude, knowledge and non-technical skills, but also objectively changes behavior. A onetime CRM-training is insufficient to change safety culture.

Additionally, the study confirms the possibility of a generic training for acute specialty areas. Between the three specialty areas (OB, ER, ICU) only limited differences were detected.

ACKNOWLEDGEMENTS

We would like to thank all participating acute care hospitals, professional associations and especially the healthcare professionals who participated in the study. Additionally, we would like to thank Margo Wouters and Marjan Goossens for support during data collection and transcription.

DECLARATIONS OF INTEREST

None.

REFERENCES

European Medicines Agency (2016). *Guideline for good clinical practice E6(R2) - step 5*. Retrieved from UK: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5_en.pdf

Institute of Medicine Committee on Quality of Health Care in America (2000). *To Err is Human: Building a Safer Health System*. Washington (DC): National Academies Press (US).

Carron, P. N., Trueb, L., & Yersin, B. (2011). High-fidelity simulation in the nonmedical domain: practices and potential transferable competencies for the medical field. *Adv Med Educ Pract, 2*, 149-155. doi:10.2147/AMEP.S19447

Chen, W., Iqbal, U., & Li, Y. J. (2017). Improving quality of care through evaluating potentially preventable events and crew resource management implementation. *Int J Qual Health Care, 29*(6), 751. doi:10.1093/intqhc/mzx127

Crofts, J. F., Bartlett, C., Ellis, D., Winter, C., Donald, F., Hunt, L. P., & Draycott, T. J. (2008). Patient-actor perception of care: a comparison of obstetric emergency training using manikins and patient-actors. *Quality and Safety in Health Care, 17*(1), 20-24. doi:10.1136/qshc.2006.021873

Flin, R., O'Connor, P., & Crichton, M. (2008). *Safety at the sharp end: a guide to non-technical skills*. Surrey: Ashgate Publishing Limited.

Franck, E., Roes, L., De Schepper, S., & Timmermans, O. (2018). Team resource management and quality of care. In *The Organizational Context of Nursing Practice* (pp. 217-237): Springer.

Fransen, A. F., van de Ven, J., Merién, A. E., de Wit-Zuurendonk, L. D., Houterman, S., Mol, B. W., & Oei, S. G. (2012). Effect of obstetric team training on team performance and medical technical skills: a randomised controlled trial. *BJOG, 119*, 1387-1393.

Fransen, A. F., van de Ven, J., Schuit, E., van Tetering, A. A. C., Mol, B. W., & Oei, S. G. (2017). Simulation-based team training for multi-professional obstetric care teams to improve patient outcome: a multicentre, cluster randomised controlled trial. *BJOG: An International Journal of Obstetrics & Gynaecology, 124*(4), 641-650. doi:10.1111/1471-0528.14369

Grogan, E. L., & et al. (2004). The impact of aviation-based teamwork training on the attitudes of health care professionals. *J Am Coll Surg, 199*(6), 843-848.

Gross, B., Rusin, L., Kiesewetter, J., Zottmann, J. M., Fischer, M. R., Prückner, S., & Zech, A. (2019). Crew resource management training in healthcare: a systematic review of intervention design, training conditions and evaluation. *BMJ Open, 9*(2), e025247. doi:10.1136/bmjopen-2018-025247

Guise, J. M., Deering, S., Kanki, B., Osterweil, P., Li, H., Mori, M., & Lowe, N. K. (2008). Validation of a Tool to Measure and Promote Clinical Teamwork. *Simulation in Healthcare, 3*(4).

Guise, J. M., & Segel, S. (2008). Teamwork in obstetric critical care. *Best Pract Res Clin Obstet Gynaecol, 22*(5), 937-951.

Haller, G., Garnerin, P., Morales, M.-A., Pfister, R., Berner, M., Irion, O., . . . Kern, C. (2008). Effect of crew resource management training in a multidisciplinary obstetrical setting. *International Journal for Quality in Health Care, 20*(4), 254-263. doi:10.1093/intqhc/mzn018

Herzberg, S., Hansen, M., Schoonover, A., Skarica, B., McNulty, J., Harrod, T., . . . Guise, J. M. (2019). Association between measured teamwork and medical errors: an observational study of prehospital care in the USA. *BMJ Open, 9*(10), e025314. doi:10.1136/bmjopen-2018-025314

Higham, H., & Baxendale, B. (2017). To err is human: use of simulation to enhance training and patient safety in anaesthesia. *Br J Anaesth, 119*(suppl_1), i106-i114. doi:10.1093/bja/aex302

Hull, L., Arora, S., Aggarwal, R., Darzi, A., Vincent, C., & Sevdalis, N. (2012). The impact of nontechnical skills on technical performance in surgery: a systematic review. *J Am Coll Surg, 214*(2), 214-230. doi:10.1016/j.jamcollsurg.2011.10.016

IBM. (2019). IBM SPSS Statistics for Windows (Version 26.0). Armonk, New York: IBM Corp.

James, J. T. (2013). A new, evidence-based estimate of patient harms associated with hospital care. *J Patient Saf, 9*(3), 122-128. doi:10.1097/PTS.0b013e3182948a69

Joint Commission International. (2014). *Joint Commission International Accreditation Standards voor Hospitals*. Retrieved from

Kao, L., & Thomas, E. (2008). Navigating Towards Improved Surgical Safety Using Aviation-Based Strategies. *J Surg Res, 145*, 327-335.

Kirkpatrick, D. L. (1979). Techniques for evaluating training programs. *Training and development journal*.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2005). *Evaluating Training Programs: The Four Levels*. Williston, VT, USA: Berrett-Koehler Publishers.

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation*: Association for Talent Development.

Kohn, L., Corrigan, J., & Donaldson, M. (2000). *To err is human: building a safer health system*. Retrieved from Washington (DC):

Leonard, M., Graham, S., & Bonacum, D. (2004). The human factor: the critical importance of effective teamwork and communication in providing safe care. *BMJ Quality & Safety, 13*(suppl 1), i85-i90.

Lingard, L., Espin, S., Whyte, S., Regehr, G., Reznick, R., Bohnen, J., . . . Grober, E. (2004). Communication failures in the operating room: an observational classification of recurrent types and effects. *Qual Saf Health Care, 13*(5), 330-334.

Manser, T. (2009). Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. *Acta Anaesthesiol Scand, 53*(2), 143-151.

Murphy, K. (2006, 31 Oktober). What Pilots Can Teach Hospitals About Patient Safety. *The New York Times*.

O'Dea, A., O'Connor, P., & Keogh, I. (2014). A meta-analysis of the effectiveness of crew resource management training in acute care domains. *Postgraduate Medical Journal, 90*(1070), 699-708. doi:10.1136/postgradmedj-2014-132800

Qualtrics. (2015). Qualtrics. Provo, Utah, USA. Retrieved from https://www.qualtrics.com

Rall, M., & Dieckmann, P. (2005). Crisis Resource Management to Improve Patient Safety. *Euroanesthesia*, 107-112.

Sexton, J. B., et al. (2006). The Safety Attitudes Questionnaire: psychometric properties benchmarking data, and emerging research. *BMC Health Serv Res, 6*, 44.

Siassakos, D., Crofts, J. F., Winter, C., Weiner, C. P., & Draycott, T. J. (2009). The active components of effective training in obstetric emergencies. *BJOG: An International Journal of Obstetrics & Gynaecology, 116*(8), 1028-1032. doi:10.1111/j.1471-0528.2009.02178.x

Siassakos, D., Fox, R., Bristowe, K., Angouri, J., Hambly, H., Robson, L., & Draycott, T. J. (2013). What makes maternity teams effective and safe? Lessons from a series of research on teamwork, leadership and team training. *Acta Obstetricia et Gynecologica Scandinavica, 92*(11), 1239-1243. doi:10.1111/aogs.12248

Sundar, E., Sundar, S., Pawloswki, J., Blum, R., Feinstein, D., & Pratt, S. (2007). Crew Resource Management and Team Training. *Anesthesiol Clin, 25*, 283-300.

Tamkin, P., Yarnall, J., & Kerrin, M. (2002). *Kirkpatrick and Beyond: A review of models of training evaluation*: Institute for Employment Studies Brighton, England.

Willems, L. M., Kurka, N., Bohmann, F., Rostek, P., & Pfeilschifter, W. (2019). Tools for your stroke team: adapting crew-resource management for acute stroke care. *Pract Neurol, 19*(1), 36-42. doi:10.1136/practneurol-2018-001966

FIGURE LEGEND

Figure 1: Score and perceived usefulness of the theoretical part, the simulation-based part and the entire training program (N=221)

Figure 2: Overall score on Clinical Teamwork Scale one month before and one month after the intervention based on observations of unexpected simulations (N=12)