



Postprint Version	1.0
Journal website	http://qualitysafety.bmj.com/content/early/2016/02/03/bmjqs-2015-003994.long
Pubmed link	http://www.ncbi.nlm.nih.gov/pubmed/26843412
DOI	10.1136/bmjqs-2015-003994

This is a NIVEL certified Post Print, more info at <http://www.nivel.eu>

Crew resource management training in the intensive care unit. A multisite controlled before–after study

PETER F KEMPER¹, MARTINE DE BRUIJNE¹, CATHY VAN DYCK², RALPH L SO³, PETER TANGKAU⁴, CORDULA WAGNER^{1,5}

¹Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands

²Faculty of Social Sciences, Department of Organisational Sciences, VU University, Amsterdam, The Netherlands

³Department of Intensive Care, Albert Schweitzer Ziekenhuis, Dordrecht, The Netherlands

⁴Department of Intensive Care, Reinier de Graaf Gasthuis, Delft, The Netherlands

⁵The Netherlands Institute of Health Services Research (NIVEL), Utrecht, The Netherlands

ABSTRACT

Introduction There is a growing awareness today that adverse events in the intensive care unit (ICU) are more often caused by problems related to non-technical skills than by a lack of technical, or clinical, expertise. Team training, such as crew resource management (CRM), aims to improve these non-technical skills. The present study evaluated the effectiveness of CRM in the ICU.

Methods Six ICUs participated in a paired controlled trial, with one pretest and two post-test measurements (after 3 and 12 months). Three ICUs received CRM training and were compared with a matched control unit. The 2-day classroom-based training was delivered to multidisciplinary groups (ie, ICU physicians, nurses, managers). All levels of Kirkpatrick's evaluation framework were assessed using a mixed method design, including questionnaires, observations and routinely administered patient outcome data.

Results Level I—reaction: participants were very positive directly after the training. Level II—learning: attitudes towards behaviour aimed at optimising situational awareness were relatively high at baseline and remained stable. Level III—behaviour: self-reported behaviour aimed at optimising situational awareness improved in the intervention group. No changes were found in observed explicit professional oral communication. Level IV—organisation: patient outcomes were unaffected. Error management culture and job satisfaction improved in the intervention group. Patient safety culture improved in both control and intervention units.

Conclusions We can conclude that CRM, as delivered in the present study, does not change behaviour or patient outcomes by itself, yet changes how participants think about errors and risks. This indicates that CRM requires a combination with other initiatives in order to improve clinical outcomes.

INTRODUCTION

Adverse events are a serious cause of harm and even deaths in the intensive care unit (ICU).^{1–3} There is a growing awareness that these adverse events are more often caused by problems with non-technical skills than from a lack of technical expertise.^{4,5} Non-technical skills are ‘the cognitive, social and personal resource skills that complement technical skills and contribute to safe and efficient task performance’.⁶ Examples of non-technical skills are task management, teamwork, situation awareness and leadership.^{4,7}

Non-technical skills are of particular importance in the ICU, where complex care for vulnerable patients is provided by multiple professionals.⁸ ICU teams have to respond 24/7 to acute situations in changing multidisciplinary compositions, under high levels of time pressure and with high stake outcomes. Under these circumstances, the chances of errors occurring increase,⁹ a fact that is reflected in a high adverse event ratio in the ICU (9.4%) compared with the average hospital ratio (5.4%).²

Team training, such as crew resource management (CRM), is a method being increasingly applied in healthcare in order to improve patient safety by applying non-technical skills, leading to an optimal use of all resources in the team and their environment.^{10,11} CRM aims to increase understanding of how certain threats and risks may lead to unsafe behaviour, as well as providing tools and training non-technical skills for responding to demanding and complex situations.

A review of Rabøl *et al*¹² shows that classroom-based CRM training consistently leads to positive first reactions and a change in attitude that favours CRM principles.^{13,14} However, the results are less conclusive with regard to the question of whether CRM changes behaviour. For instance, McCulloch *et al*¹⁵ found an increase in the use of non-technical skills for nurses, but not for anaesthetists and surgeons. There are several reasons for these inconsistent findings. Most studies have a relatively short follow-up period, they lack a control group and/or rely on self-reported outcomes only. Therefore, the precise effects of CRM are still unclear.

Building on the shoulders of our predecessors, we conducted an evaluation in which we tried to overcome some of these shortcomings by including a control arm, independent measurement of behaviour and 12-months follow-up. We used the Kirkpatrick evaluation framework for training programmes to assess the effectiveness.¹⁶ This framework distinguishes four levels of evaluation: (1) reaction, (2) learning, (3) behavioural change and (4) organisational impact. This framework allows for a systematic and multilevel evaluation, while at the same time organising it in a simple and understandable manner.¹⁷ Consequently, our central research

question was: what is the effectiveness of a classroom-based CRM training in the ICU, in which all staff receive training, on all levels of Kirkpatrick's framework compared with matched control ICUs without training?

MATERIALS AND METHODS

We developed a mixed method design aimed at assessing all levels of the Kirkpatrick framework, including questionnaires, direct observations and analysis of routine administrative data (see table 1). A mixed method design is recommended for capturing the effects of diffuse and complex innovations, such as CRM, as it triangulates the effect of an intervention and increases the valid interpretation of the results.¹⁸

[TABLE 1]

The design of the present study was published in detail in advance.¹⁹ A summary of relevant aspects of the design is given below.

Design and setting

The present study applied a multicentre controlled before and after design. By using a control arm and a premeasurement and postmeasurement, the possible effects of patient safety initiatives in the ICU instigated without CRM could be controlled for.

Three pairs of comparable ICUs participated in the present study. For each pair, one ICU received the CRM training and one ICU acted as a control group. All ICUs were part of non-academic teaching hospitals (a total of 806 beds on average) in the Netherlands. All specialties were available in the hospitals except cardiac surgery and neurosurgery, which were available in nearby hospitals. The ICUs were mixed medical surgical, intensivist-led and contained 10–17 beds and 63–95 unique employees. This type of ICU is large enough to form an independent unit, yet small enough for the whole staff to be trained. Intervention ICUs needed to comply with formal criteria, like financial and organisational arrangements. Control units were matched to a specific intervention unit based on the number of beds and ICU physicians (fulltime equivalents, FTEs), the perception of patient safety, frequency of event reporting and an interview with the two of the authors. A detailed description of the matching procedure can be found in Kemper *et al.*¹⁹

The workforce of the ICUs includes IC physicians, residents, nurses, student nurses and managers, all of whom have been included in the current study. All ICU staff members (n=474) were invited to participate in the study, see table 1 for the response rate.

The present study comprised one premeasurement and two follow-up measurements. The pretest was conducted in a period of 7–9 weeks before the training. The first follow-up measurement was carried out 3 months after the training and the second measurement followed 10 months after the training.

The study was approved by the Ethical Committee of the VU University Medical Center and is in accordance with Dutch privacy regulations. The trial is registered in the Dutch Trial Registration record NTR1976. Participation was confidential and all data collected were stored entirely anonymously.

CRM training

The CRM training was classroom based and consisted of an education session of two consecutive days from 09:00 to 17:00 in groups of 15 participants. A commercial vendor of CRM, QST Safe Skies, was contracted to deliver the training. This vendor has much experience with classroom-based CRM trainings in the aviation sector, as well as in healthcare. This particular vendor was chosen because of the focus on a bottom-up approach and the usage of the input of the participants in line with the adult learning principles.²⁰ Importantly, this vendor allowed an independent study alongside the training and the required time and costs were feasible for the participating ICUs.

All members of the IC staff (n=193) were trained during office hours in multidisciplinary groups. During the training course, new ideas for safety initiatives were continuously gathered and registered in the form of concrete plans of action. An example of an action on the team level is to explicitly appoint a coordinator of the day in order to improve and clarify leadership, teamwork and communication. The plans of action differed each training and for each ICU. After the training, all ICUs started implementing these newly developed initiatives. This implementation process is described elsewhere.²¹ Part of the training was that the CRM instructors offered their help as consultant for 1 day after the ICU was trained. It was up to the change team how to use this help. An extensive description of the CRM training as applied in the present study was published in advance (box 1).¹⁹

[Box 1]

Measurements

The measurements were split up according to the four levels of the Kirkpatrick evaluation framework.¹⁶ Table 1 shows an overview of the various instruments that each level comprises and when the measurement was conducted. All applied instruments were previously validated. All measurements were administered simultaneously in each pair of the intervention and control unit. The questionnaires regarding reaction to and evaluation of the training were administered directly and 3 months after the training, respectively. When applicable, the Cronbach's α of the dimensions of the questionnaires is included in the overview of the results within the paper or in the online supplementary appendix, as an indicator of the reliability of the scale.

Level I: reaction

In line with the work of Alliger *et al*,¹⁷ we divided the level of reaction into an affective reaction (End-of-Course Critique, ECC) and the perception of the utility of the training programme (evaluation questionnaire).

End-of-Course Critique

The ECC of Grogan *et al*¹⁴ was used for the present study to assess the affective reaction immediately after the training. Grogan's ECC was adapted to fit the present study, for instance by asking about the topics that were discussed in the current CRM training. All questions are shown in the online supplementary appendix table A1.

Evaluation questionnaire

This questionnaire was newly developed and used to assess the extent to which the training altered the participant's awareness regarding CRM topics in their daily work. The questions were straightforward formulated covering most topics that were discussed during the training, for instance 'I'm more aware than before of personal factors influencing my performance'. All questions are shown in the online supplementary appendix table A2.

Level II: learning from the training programme

As a measure of learning, we studied whether the participants changed the way they thought about relevant aspects regarding the CRM training. In other words, whether their attitude was changed. We focused on situational awareness (SA), as this can be seen as the central element of CRM.

SafeTeamA questionnaire: SA attitudes

This questionnaire, developed and validated alongside the present study,²⁶ was used to assess the attitude towards behaviour that optimises SA. Endsley²² defines SA as: 'The perception of elements in the environment (...), the comprehension of their meaning, and the projection of their status in the near future'. SA is expected to optimise through applying specific tactics that stem from non-technical skills, such as speaking up and the sharing of information. The SafeTeamA questionnaire assesses to what extent a participant perceives these tactics to be important. This is measured with one scale comprising 11 items.

Level III: behavioural change

Behavioural change reflects the level of transfer of what is learned during the training course and applying this to the work setting. We assessed the extent of this transfer by asking the participants about their behaviour and through observing them.

SafeTeamB questionnaire: enacted SA tactics

SafeTeamB follows the same rationale as the SafeTeamA questionnaire,²⁶ but measures the self-reported appliance of the tactics that optimises SA, rather than attitudes. Furthermore, it comprises two scales, one measuring individual behaviour (seven items) and one measuring team behaviour (five items).

EPOC: observation of non-technical skills

Direct observations were used to determine the use of non-technical skills by the IC staff, using the EPOC observation tool (EPOC).²⁷ Observations were carried out with consent of the observed person. The EPOC assesses the amount of explicit professional oral communication during an observation of 30 min by a trained observer, during daily practice, taking several control variables into account (eg, workload). The EPOC has been successfully applied in another study.²⁸

The EPOC does not require clinical knowledge of the observer. For each measurement period, a group of four observers conducted the observations. Observers were students that received a 2-day training course, in which they learned the definitions of the verbal behaviours, practised with the observation tool and were familiarised with common observation biases, such as the Halo effect. Unfortunately, it was not possible to recruit the same observers during both measurements. There was a good interobserver reliability as the intraclass correlation was mostly >0.60 .²⁷ More information about the EPOC, its validity and reliability and the training is described by Kemper *et al.*²⁷

Level IV: organisation

The level of organisation, or ‘results’, concerns the impact of the training programme on the organisation. Outcomes on this level can be very diverse and may take time before changes become noticeable.²⁹ We studied the patient outcomes, parts of the organisational culture and job satisfaction on this level.

Patient outcomes

Length of stay, readmissions within 24 h and ICU mortality were used as indicators of the quality of care, using current registration standards from the Dutch National Intensive Care Evaluation (NICE) study.³⁰ The APACHE IV³¹ score was calculated in the present study to adjust for differences in the patient mix. In addition to the APACHE IV, other confounders were taken into consideration as well, notably gender; use of mechanical ventilation with admission; use of mechanical ventilation within the first 24 h of admission; whether the admission was scheduled or not.

Patient safety culture

The COMPaZ questionnaire³² was used to measure the patient safety culture in the ICU. The COMPaZ is the translated and validated Dutch version of the Hospital Survey on Patient Safety Culture³³ and successfully applied in various studies.^{34, 35}

Error culture

The Error Culture Questionnaire (ECQ), developed and validated by van Dyck *et al*^{36,37} was used to assess shared attitudes towards, and common responses to, errors at the unit level. The ECQ has been successfully used in previous research.^{37,38}

Job satisfaction and affective commitment

Job satisfaction was measured using three of six scales of the Dutch translation³⁹ of the job satisfaction dimension of the Occupational Stress Inventory,⁴⁰ that is, satisfaction with (1) the job, (2) the organisational design and structure and (3) the organisational processes. One additional item was added concerning how satisfied the participants were with their job.

Affective commitment was assessed using the Dutch translation⁴¹ of the affective subscale of the three-component conceptualisation of organisational commitment.⁴²

Data analysis

The level of reaction was analysed only descriptively, as no comparison could be made with either a control group or a premeasurement. All other outcomes were analysed using different analyses for the observations, the patient outcomes and the questionnaires, as described below. However, all analyses had in common that the main outcome was the interaction component, indicating whether the difference between the premeasurement and postmeasurement of the intervention group was significantly deviant from the difference observed in the control group.

Observations

Since participants were observed multiple times, a linear mixed model was used to adjust for possible clustering of data within a participant. The crude model comprised a measurement variable, an experimental group component and the interaction component between these two variables. The crude model was tested for possible influence of several confounders, such as the occupation of the observed person. These variables were included when they significantly improved the model, expressed as a significant decrease in the log-likelihood ratio.

Patient outcomes

Linear and logistic regression analyses were used to determine whether differences between the intervention and control group as a result of the CRM training were statistically significant. The crude model consisted of the measurement variable, the experimental group variable and the interaction component between these two variables. The analyses of the patient outcomes comprised two steps. First, we analysed to what extent confounders influenced the outcome variable (the interaction component). If the regression coefficient of the interaction component of the crude model was altered by more than 10% after including the supposed confounder in the model, then this confounder was included in the final model of step 2. The second

step consisted of a multivariate analysis with the crude model and the influential confounders of the first step.

Questionnaires

Differences between experimental groups and measurements were analysed with an analysis of variance for repeated measures. This method has the advantage of testing the within-participant difference, thereby partitioning out variability due to individual differences.

RESULTS

Level I: reaction to the training programme

End-of-course questionnaire

Overall, the respondents were very positive about the training. The themes were rated as highly relevant and useful ($M=4.47$; $SD=0.45$). Participants indicated that the topics were new and informative ($M=3.81$; $SD=0.68$) and that due to the training, they had developed new ideas and topics that they could directly apply to their work ($M=4.31$; $SD=0.61$). Participants would definitely recommend the training to colleagues ($M=4.50$; $SD=0.62$). An overview of the results for each item is shown in the online supplementary appendix table A1.

Evaluation questionnaire

This questionnaire demonstrates that participants perceived themselves as having changed 'slightly' to 'reasonably' 3 months after the training with regard to the assessed aspects of the training. The overall mean of all 13 items is 2.71 ($SD=0.69$) on a 5-point scale. An overview of the results for each item is shown in the online supplementary appendix table A2.

Level II: learning from the training programme

SafeTeamA questionnaire

The results showed that participants perceive that it is important to have an optimal SA even before CRM training. This is reflected in a high mean score on the SafeTeamA questionnaire in both the intervention ($M=4.19$; $SD=0.37$) and control ($M=4.16$; $SD=0.39$) group during the baseline measurement. The intervention group demonstrated a small increase after the training ($M=4.21$; $SD=0.40$), whereas this decreased in the control group ($M=4.12$; $SD=0.41$). However, this change was not significantly different ($F=1.68$; $p=0.20$). The results are shown in the online supplementary appendix table A3.

Level III: behavioural change as a result of the training

SafeTeamB questionnaire

The results of the SafeTeamB questionnaire, depicted in the online supplementary appendix table A3, showed that participants that had followed the training perceived more enacted SA tactics in their individual behaviour, as well as in their team behaviour. The increase of individual self-reported behaviour in the intervention group (pre: M=2.92; SD=0.55; post: M=3.21; SD=0.56) was significantly different ($F=8.02$; $p<0.01$) from the change in the control group (pre: M=3.08; SD=0.50; post: M=3.19; SD=0.50). Likewise, the increase of team self-reported behaviour in the intervention group (pre: M=3.46; SD=0.50; post: M=3.70; SD=0.45) was significantly different ($F=16.61$; $p<0.01$) from the change in the control group (pre: M=3.69; SD=0.47; post: M=3.70; SD=0.51).

EPOC: observation of non-technical skills

We conducted 1006 observations in total. In the premeasurement, we observed 82% of the staff at least once and in the postmeasurement, this percentage was 90%. The descriptive results (median and 25th and 75th percentiles) as well as the regression coefficient of the interaction component are shown in table 2. The amount of explicit professional oral communication decreased after the training in both the intervention group and the control group. No statistically significant differences in the change of explicit communication were found between the intervention group and control group.

[TABLE 2]

Level IV: organisational impact of the training

Patient outcomes

During the premeasurement and postmeasurement, 8027 patients were admitted to the participating ICUs. The results presented in table 3 show no statistically significant deviations between the intervention group and the control group, when comparing the difference between the premeasurement and postmeasurement (ie, the interaction effect). In other words, there are no changes in outcomes that can be attributed to the CRM training.

[TABLE 3]

Patient safety culture

The results regarding the patient safety culture in the ICU generally showed a positive trend. Both intervention group (before: M=3.44, SD=0.29; after: M=3.61, SD=0.32) and control group (before: M=3.55, SD=0.27; after: M=3.67, SD=0.33) showed an increase in the sum score of all dimensions. However, regression analysis did not show a statistically significant overall effect of CRM on culture. In the intervention group, two dimensions changed significantly. ‘Supervisor/manager expectations and actions promoting patient safety’ was significantly influenced by CRM training ($F=7.45$, $p<0.05$). The intervention group rated this dimension more positively (before: M=3.38, SD=0.58; after: M=3.66, SD=0.47) after the training, than did the control group (before: M=3.46, SD=0.56; after: M=3.50, SD=0.58).

Furthermore, 'Adequate staffing' was rated significantly higher ($F=9.69$; $p<0.01$), as the intervention group showed an increase (before: $M=3.45$, $SD=0.79$; after: $M=3.85$, $SD=0.54$) while the control group did not change (before: $M=3.86$, $SD=0.42$; after: $M=3.98$, $SD=0.46$). The other dimensions did not show changes as a result of the training. An overview of the results for each dimension is shown in the online supplementary appendix table A4.

Error culture

The results, depicted in table 4, showed that the four dimensions of the ECQ seem to have changed as a result of the CRM training. Participants indicated that they were reacting more socially by helping each other and communicating about errors. Additionally, participants showed more error awareness, an effect that can be predominantly attributed to more anticipation of errors. The Mastery dimension showed a significant change in the intervention group as well, indicating that participants try to overcome errors. The underlying concepts of Mastery are learning, correcting and analysing errors, with analysing showing the most change. The changes in the Aversion dimension did not reach a level of significance, although a trend can be distinguished as the p value is just above 0.05 ($p=0.06$). An overview of the results for each scale underlying the presented dimension is displayed in the online supplementary appendix table A5.

[TABLE 4]

Job satisfaction and affective commitment to the ICU

Table 5 shows that all dimensions had a significant interaction component, except for the job satisfaction in general. This means that the changes of the intervention group between the premeasurements and postmeasurements are significantly different from those of the control group. In other words, the positive changes in job satisfaction and affective commitment only occur in the intervention group and not in the control group.

[TABLE 5]

DISCUSSION

The present study aimed at providing a multilayered evaluation to determine the effectiveness of classroom-based CRM training on all levels of the Kirkpatrick evaluation framework. The first reaction to the training was positive, with self-reported enacted situation awareness tactics, error culture and job satisfaction improving significantly in the ICUs that received CRM training, whereas the ICUs that did not receive CRM training showed no changes. There was no change in explicit professional communication, neither did patient outcomes improve. This is an indication that improvements at the sharp end require a further implementation of CRM initiatives that go beyond the 2 days of training and a year-long course of activities of the CRM change teams after the training. The results are summarised in figure 1.

[FIGURE 1]

The framework of Kirkpatrick was used to assess four different levels of effect of CRM. Measurements at the first level show that participants were very positive directly after the training. This is in line with previous research.^{14,43} However, 3 months after CRM, participants attributed only a moderate change in their behaviour as a result of CRM. This could be a signal that participants find it difficult to translate CRM to their daily work. Anchoring CRM in the organisation or continuous training could help overcome this.

Attitudes towards tactics to optimise SA, measured at the second level of the Kirkpatrick framework, were stable, despite a small and non-significant increase in the intervention group. A possible explanation for this result is that the attitudes were reasonably high at baseline, in both the intervention group and control group, making it difficult to achieve an even higher attitude.

Self-reported behaviour regarding tactics to optimise SA did improve as a result of CRM training. This indicates that CRM empowers participants to behave in a way that corresponds with their high appreciation expressed in the attitude assessment, whereas members of the control units share the high appreciation but do not act on it. Other research also shows that a change in attitude is not a prerequisite for behavioural change. McCulloch *et al*¹⁵ also found that personnel in the operating theatre performing a laparoscopic cholecystectomy showed an increased use of non-technical skills, while the attitude towards teamwork climate remained the same.

At the third level of the Kirkpatrick framework, we expected the amount of explicit professional oral communication to increase as a result of CRM training, assuming that the amount of professional communication was limited in advance. Research in the emergency department (ED) confirmed this hypothesis, as the amount of explicit communication increased after the ED staff received the training. This effect was, however, not found in the present study. This might be attributed to the fact that the amount of professional communication is almost 5 times higher in the ICU compared with the ED.²⁷ In departments such as the ICU, with continuous teamwork for each patient, the amount of professional communication is relatively high. Improving focused communication and standardising processes, may reduce, rather than increase, the total amount of professional communication.^{44,45}

The fourth and last level of the Kirkpatrick framework comprised multiple measurements. Patient outcomes such as length of stay were not affected by the CRM training. Although multiple initiatives were developed within all ICUs, the implementation approach and focus of the initiatives differed per ICU.²¹ Without a shared focus between the ICUs on one initiative with a subsequent outcome, we can in retrospect argue that a change in one of the applied outcomes cannot be expected. It can be argued that the combined effect of multiple CRM improvement projects (ie, the development of a checklist for high-risk situations) could eventually improve patient outcomes, though this would require further implementation. Studies with a single focus within the team training across sites on briefings and debriefings have shown promising results such as decreasing delay in surgery⁴⁶ and even reducing

mortality.⁴⁷ Not finding an effect on patient outcomes is however not uncommon when evaluating a CRM training.¹² This brings Weaver and Rosen⁴⁸ to conclude their examination of reviews on the effectiveness of medical team trainings that in general team trainings are associated with improvements in patient safety outcomes, rather than being the causal precursor. This indicates that there are more factors to be considered that might influence the effectiveness of a team training.

Scores on the dimensions of the patient safety culture show an upward trend in both the intervention and control departments, with the exception of ‘expectations and actions of leaders and supervisor regarding the promotion of patient safety’ and ‘adequate staffing’. The changes in these two organisationally rooted dimensions can be attributed to the CRM training. A possible explanation for this result could merely be the fact that CRM was organised. Organising such a training is a clear signal that patient safety is important.

The way people respond to errors—the error culture—changed in the ICUs that received CRM training. In this group, respondents perceived themselves and their colleagues as being more aware regarding errors and being more socially oriented on the occasions they occur. Furthermore, after the training, participants responded that they thought it was important to overcome errors in the future and, to a lesser extent, be less error aversive. These changes reflect the core principle of CRM of recognising and preventing errors, and mitigating the consequences when errors do occur.

Job satisfaction and affective commitment increased in the intervention group. There are several possible explanations for these results. Participants get to know their colleagues better while spending 2 days with them outside the clinical context. This alone could be beneficial for job satisfaction. Additionally, during the training, all roles of the various disciplines were discussed. This stimulated mutual understanding of the role of other team members, which in turn could have increased mutual understanding and reduced possible reciprocal annoyance.

Limitations

A limitation of the present study is that we were unable to assess errors, incidences or adverse events as a measure of patient outcomes. The current registration of complications that is supported by NICE was still in its infancy, and therefore did not provide useful and reliable data. It was, however, possible to determine reliable quality indicators such as length of stay, death in the ICU and readmissions within 24 h. We expect these indicators to behave in the same way the complications would have, if measured reliably. Based on this line of reasoning, it is not likely that we would have found a reduction in incidences or adverse events as a result of the training within this time frame.

In order to broadly assess effectiveness, the present study used a mixed method approach and closely monitored the participating ICUs. This enabled us to draw conclusions from multiple instruments and perspectives and to seek corroboration in these conclusions. This process is called triangulation and is the major benefit of a

mixed method.¹⁸ The downside of this approach is that it is time-consuming, thereby limiting the number of ICUs that could be included, especially because the current study is bound to a maximum duration of 3 years as a prerequisite of the governmental fund that made this study possible. This limit on the number of participating ICUs is unfavourable in terms of the statistical power of the study. The benefit, however, was that there were enough ICUs willing to participate in the present study as a control unit, which enabled a thorough selection process to take place so that the best match for the intervention units could be made.

Finally, it can be argued that the present study faced a reporting bias, in the sense that participants who receive CRM training learn the ‘right’, or most social acceptable, answers. We expect that this bias was not present in the present study, as the measurement that is most likely to suffer this kind of bias, the attitude questionnaire, did not show any differences that could be attributed to the training. Additionally, self-reported questionnaires suffer the bias of limited self-knowledge of the respondent.⁴⁹

CONCLUSIONS

Based on the results, we can conclude that the 2-day classroom-based CRM training, as delivered in the present study, does not change behaviour or patient outcomes by itself, yet changes how participants think about errors and risks. It is plausible that when CRM is combined with other initiatives and is prolonged over a period of time it may have an effect on behaviour and patient outcomes. Participants perceive that errors can be more openly discussed and try to anticipate possible risks. It can be argued that these changes make the staff more receptive for patient safety and quality improvement initiatives, and that CRM can even be used to develop such interventions. Therefore, we recommend repeating CRM and embedding it in the organisation for the purpose of further exploiting the impetus mechanism, for instance as part of continuous medical education, and incorporating it in a more focused, multifaceted implementation strategy in order to produce changes on a behavioural and organisational level.

Footnotes

- Contributors PFK and MdB conceived the study, acquired and analysed the data and drafted the manuscript. CvD and CW helped to interpret the results and to draft the manuscript. RLS and PT participated in conceiving the design, acquiring the data and drafting the manuscript. All authors read and approved the final manuscript.
- Funding ZonMw (170992804).
- Competing interests None declared.
- Ethics approval Ethical Committee of the VU University Medical Center.
- Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- Kohn LT, Corrigan JM, Donaldson MS. *To err is human: Building a safer health care system*. Washington DC: National Academy Press, 2000.
- Zegers M, de Bruijne MC, Wagner C, et al. Adverse events and potentially preventable deaths in Dutch hospitals: results of a retrospective patient record review study. *Qual Saf Health Care* 2009;18:297–302. doi:10.1136/qshc.2007.025924
- Cook TM, Woodall N, Harper J, et al. Major complications of airway management in the UK: results of the Fourth National Audit Project of the Royal College of Anaesthetists and the Difficult Airway Society. Part 2: intensive care and emergency departments. *Br J Anaesth* 2011;106:632–42. doi:10.1093/bja/aer059
- Yule S, Flin R, Paterson-Brown S, et al. Non-technical skills for surgeons in the operating room: a review of the literature. *Surgery* 2006;139:140–9. doi:10.1016/j.surg.2005.06.017
- Wolff AM, Bourke J. Reducing medical errors: a practical guide. *Med J Aust* 2000;173:247–51.
- Flin R, Patey R, Glavin R, et al. Anaesthetists' non-technical skills. *Br J Anaesth* 2010;105:38–44. doi:10.1093/bja/aeq134
- Reader T, Flin R, Lauche K, et al. Non-technical skills in the intensive care unit. *Br J Anaesth* 2006;96:551–9. doi:10.1093/bja/ael067
- Haerkens MH, Jenkins DH, van der Hoeven JG. Crew resource management in the ICU: the need for culture change. *Ann Intensive Care* 2012;2:39. doi:10.1186/2110-5820-2-39
- Donchin Y, Gopher D, Olin M, et al. A look into the nature and causes of human errors in the intensive care unit. 1995. *Qual Saf Health Care* 2003;12:143–7. doi:10.1136/qhc.12.2.143
- Flin R, Maran N. Identifying and training non-technical skills for teams in acute medicine. *Qual Saf Health Care* 2004;13(Suppl 1):i80–4. doi:10.1136/qshc.2004.009993
- Helmreich RL. On error management: lessons from aviation. *BMJ* 2000;320:781–5. doi:10.1136/bmj.320.7237.781
- Rabøl LI, Østergaard D, Mogensen T. Outcomes of classroom-based team training interventions for multiprofessional hospital staff. A systematic review. *Qual Saf Health Care* 2010;19:e27. doi:10.1136/qshc.2008.029967
- France DJ, Stiles R, Gaffney EA, et al. Crew resource management training—clinicians' reactions and attitudes. *AORN J* 2005;82:213–24. doi:10.1016/S0001-2092(06)60313-X
- Grogan EL, Stiles RA, France DJ, et al. The impact of aviation-based teamwork training on the attitudes of health-care professionals. *J Am Coll Surg* 2004; 199:843–8. doi:10.1016/j.jamcollsurg.2004.08.021
- McCulloch P, Mishra A, Handa A, et al. The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre. *Qual Saf Health Care* 2009;18:109–15. doi:10.1136/qshc.2008.032045
- Kirkpatrick DL, Kirkpatrick JD. *Evaluating Training Programs. The Four Levels*. San Francisco: Berrett-Koehler Publishers Inc, 2006.
- Alliger GM, Tannenbaum SI, Bennett W, et al. A meta-analysis of the relations among training criteria. *Personnel Psychol* 1997;50:341–58. doi:10.1111/j.1744-6570.1997.tb00911.x
- Brown C, Hofer T, Johal A, et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 4. One size does not fit all. *Qual Saf Health Care* 2008;17:178–81. doi:10.1136/qshc.2007.023663
- Kemper PF, de Bruijne MV, van Dyck C, et al. Effectiveness of classroom based crew resource management training in the intensive care unit: study design of a controlled trial. *BMC Health Serv Res* 2011;11:304. doi:10.1186/1472-6963-11-304
- Kolb DA. *Experiential learning: experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall, 1984.
- Kemper PF, van Dyck DC, Wagner C, et al. Implementation of crew resource management: a qualitative study in 3 intensive care units. *J Patient Saf* 2014.
- Endsley MR, Garland DJ, eds Endsley MR. Theoretical underpinnings of situational awareness: a critical review. In: Endsley MR, Garland DJ, eds. *Situation awareness analysis and measurement*. Mahwah: Lawrence Erlbaum Associates, 2000;3–32.
- Reason J. *Human error: models and management*. *West J Med* 2000;172:393–6. doi:10.1136/ewjm.172.6.393

- Schulz von Thun F. *Hoe bedoelt u? (What do you mean?)*. Groningen: Noordhoff Uitgevers BV, 2003.
- Janis I. *Victims of groupthink: a psychological study of Foreign-policy decisions and Fiascoes*. Boston: Houghton Mifflin, 1972.
- Hertel G, Binnewies C, Krumm S, et al. Kemper PF, Geraets SJA, De Bruijne MC, et al. *Assessment of situational awareness. Team resources brought in action to enhance patient safety at the ICU*. In: Hertel G, Binnewies C, Krumm S, et al., eds. *Abstract proceedings of the 16th EAWOP Congress 2013*, 2013;751–2.
- Kemper PF, van Noord I, de Bruijne MC, et al. *Development and reliability of the explicit professional oral communication observation tool to quantify the use of non-technical skills in healthcare*. *BMJ Qual Saf* 2013;22:586–95. doi:10.1136/bmjqs-2012-001451
- Verbeek-van Noord I, de Bruijne MC, Twisk JW, et al. *More explicit communication after classroom-based crew resource management training: results of a pragmatic trial*. *J Eval Clin Pract* 2015;21:137–44. doi:10.1111/jep.12261
- Salas E, Burke CS, Bowers CA, et al. *Team training in the skies: does crew resource management (CRM) training work?* *Hum Factors* 2001;43:641–74. doi:10.1518/001872001775870386
- Arts D, de Keizer N, Scheffer GJ, et al. *Quality of data collected for severity of illness scores in the Dutch National Intensive Care Evaluation (NICE) registry*. *Intensive Care Med* 2002;28:656–9. doi:10.1007/s00134-002-1272-z
- Brinkman S, Bakhshi-Raiez F, Bu-Hanna A, et al. *External validation of acute physiology and chronic health evaluation IV in dutch intensive care units and comparison with acute physiology and chronic health evaluation II and simplified acute physiology score II*. *J Crit Care* 2010;26: 105.e11–18.
- Smits M, Christiaans-Dingelhoff I, Wagner C, et al. *The psychometric properties of the 'Hospital Survey on Patient Safety Culture' in Dutch hospitals*. *BMC Health Serv Res* 2008;8:230. doi:10.1186/1472-6963-8-230
- Sorra JS, Nieva VF. *Hospital Survey on Patient Safety Culture*. Rockville, MD: Agency for Healthcare Research and Quality, 2004.
- Marshall DA, Manus DA. *A team training program using human factors to enhance patient safety*. *AORN* 2007;86:994–1011. doi:10.1016/j.aorn.2007.11.026
- van Noord I, de Bruijne MC, Twisk JW. *The relationship between patient safety culture and the implementation of organizational patient safety defences at emergency departments*. *Int J Qual Health Care* 2010;22:162–9. doi:10.1093/intqhc/mzq013
- van Dyck C. *Putting errors to good use: Error management culture in organizations*. Amsterdam: University of Amsterdam, 2000.
- van Dyck C, Frese M, Baer M, et al. *Organizational error management culture and its impact on performance: a two-study replication*. *J Appl Psychol* 2005;90: 1228–40. doi:10.1037/0021-9010.90.6.1228
- Hofmann DA, Mark B. *An investigation of the relationship between safety climate and medication errors as well as other nurse and patient outcomes*. *Personnel Psychology* 2006;59:847–69. doi:10.1111/j.1744-6570.2006.00056.x
- Broers P, Evers A, Cooper CL. *Differences in Occupational Stress in three European Countries*. *Int J Stress Manag* 1995;2:171–80. doi:10.1007/BF01681847
- Cooper CL, Marshall J. *Occupational sources of stress—review of literature relating to coronary heart-disease and mental ill health*. *J Occup Psychol* 1976;49:11–28. doi:10.1111/j.2044-8325.1976.tb00325.x
- De Gilder D, Van den Heuvel H, Ellemers N. *A three component model of organizational commitment*. *Behav Organ (Gedrag en Organisatie)* 1997;10:95–106.
- Allen NJ, Meyer JP. *The Measurement and Antecedents of Affective, Continuance and Normative Commitment to the Organization*. *J Occup Psychol* 1990;63:1–18. doi:10.1111/j.2044-8325.1990.tb00506.x
- Haller G, Garnerin P, Morales MA, et al. *Effect of crew resource management training in a multidisciplinary obstetrical setting*. *Int J Qual Health Care* 2008;20: 254–63. doi:10.1093/intqhc/mzn018
- Marshall S, Harrison J, Flanagan B. *The teaching of a structured tool improves the clarity and content of interprofessional clinical communication*. *Qual Saf Health Care* 2009;18:137–40. doi:10.1136/qshc.2007.025247

Catchpole KR, de Leval MR, McEwan A, et al. *Patient handover from surgery to intensive care: using Formula 1 pit-stop and aviation models to improve safety and quality. Paediatr Anaesth* 2007;17:470–8. doi:10.1111/j.1460-9592.2006.02239.x

Wolf FA, Way LW, Stewart L. *The efficacy of medical team training: improved team performance and decreased operating room delays: a detailed analysis of 4863 cases. Ann Surg* 2010;252:477–83.

Neily J, Mills PD, Young-Xu Y, et al. *Association between implementation of a medical team training program and surgical mortality. JAMA* 2010;304:1693–700. doi:10.1001/jama.2010.1506

Shekell PG, Wachter RM, Pronovost PJ, et al. Weaver SJ, Rosen MA. *Chapter 40 Team-Training in health care: a brief update review. In: Shekell PG, Wachter RM, Pronovost PJ, et al., eds. Making health care safer II: an updated critical analysis of the evidence for patient safety practices. Evidence Reports/Technology Assessments. Rockville, MD: Agency for Healthcare Research and Quality (US), 2013;472–79.*

Davis DA, Mazmanian PE, Fordis M, et al. *Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA* 2006;296:1094–102.

TABLES AND FIGURES

Table 1 Study overview

Measurements per level of Kirkpatrick	Measurement properties				Data collection period and response			
	No. of dimensions	No. of items (k)	Scale	New or existing	Baseline	After the training	First follow-up	Second follow-up
I Reaction to the training programme								
End-of-Course Critique questionnaire*	n/a	14	5-Likert	Existing		100%		
Evaluation questionnaire*	n/a	13	5-Likert	New			75%	
II Learning from the training programme								
Situational awareness attitudes (SafeTeamA questionnaire)	7	43	5-Likert	New	87%			72%
III Behavioural change as a result of the training								
Enacted situational awareness tactics (SafeTeamB questionnaire)	7	39	5-Likert	New	87%			72%
EPOC: Observation of explicit professional oral communication	6	30	Count	Existing	489 observations			596 observations
IV The impact of the training on the organisation								
Routine (patient) administration data	n/a	n/a	n/a	Existing	One year before CRM: 4085 patients. One year after: 3942 patients			
Patient safety culture (COMPaz)	11	40	5-Likert	Existing	76%			72%
Error culture (error culture questionnaire)	11	47	5-Likert	Existing	87%			73%
Job satisfaction and affective commitment to the ICU	7	18	5-Likert	Existing	87%			72%

Overview of the measurements for each level of the Kirkpatrick framework with regard to their measurement properties, when they were collected and what the response was during that data collection period.

*These measurements were collected in the intervention groups only.

CRM, crew resource management; ICU, intensive care unit; n/a, not applicable.

Box 1 A specification of the main content of the crew resource management (CRM) training

The main objectives of the training sessions were to create awareness regarding the threats of suboptimal performance and ways to recognise these threats and prevent their negative consequences by optimal use of team resources. To establish this goal, the participants were educated about CRM concepts and principles, discussed their own experiences with each other and developed ready to use ideas, all in a setting of trust and openness. All topics were first introduced, then discussed and finally solutions to overcome the mentioned risks were formulated.

With situational awareness²² as a starting point to identify pitfalls and opportunities for enhancement of patient safety, several topics were discussed on an organisational, team and individual level:

Organisation	Team	Individual
▶ Human factors and CRM	▶ Situational leadership	▶ Stress management
▶ Threats and errors	▶ Communication and cooperation	▶ Personality
	▶ Groupthink	▶ Feedback and self-reflection

Various theories and models were discussed during the training, such as the Swiss cheese model,²³ the sender–receiver model,²⁴ and Groupthink.²⁵ To translate CRM concepts into ready-to-use ideas, team roles were interactively defined throughout the session and plans of action were formulated by the participants.

The content of the training is described in more detail elsewhere.¹⁹

Table 2 Explicit professional oral communication

Categories and dimensions	Intervention		Control		Interaction component β (95% CI)
	Pre (n=252)	Post (n=249)	Pre (n=237)	Post (n=268)	
EPOC total score*	41 (28 to 55)	35 (24 to 49)	35 (23 to 50.50)	30 (21 to 41.75)	1.67 (−2.30 to 5.64)
Self†	0.00 (0.00 to 0.00)	−0.09 (−25 to 0.07)			
Human interaction‡	38 (26 to 57.75)	31 (22 to 45)	34 (22 to 48)	27 (19 to 37)	1.61 (−2.13 to 5.34)
Environment§	1 (0 to 3)	3 (1 to 5)	1 (0 to 3)	2 (1 to 4)	0.18 (−0.39 to 0.75)

The descriptive results (median, 25th and 75th percentiles) and regression coefficient of the interaction component between experimental group and measurement.

More information about the categories, dimensions and underlying items are described in Kemper *et al.*²⁷

*The crude model was adjusted for: random intercept, ICU, number of people the observed person spoke to during the observation, observer, work pressure with random slope, occupation, whether or not the observed person was accompanied by a student nurse or resident, whether administration was part of the observation.

†The crude model was adjusted for: random intercept, random slope interaction component, observer.

‡The crude model was adjusted for: random intercept, random slope interaction component, ICU, number of people the observed person spoke to during the observation, observer, work pressure with random slope, occupation, whether or not the observed person was accompanied by a student nurse or resident, whether administration was part of the observation.

§The crude model was adjusted for: random intercept, ICU, number of people the observed person spoke to during the observation with a random slope, observer, work pressure with random slope, occupation, whether or not the observed person was accompanied by a student nurse or resident, whether administration was part of the observation.

ICU, intensive care unit.

Table 3 Patient outcomes

	Intervention		Control		Interaction component
	Pre (n=2549)	Post (n=2370)	Pre (n=1536)	Post (n=1572)	
Female (%)	44	44	44	44	
Male (%)	56	56	56	56	
Age (M, SD)	65 (17)	65 (16)	64 (17)	65 (16)	
APACHE IV score (M, SD)	62 (32)	65 (31)	62 (33)	63 (32)	
Mechanical ventilation with admission*	32.5%	34.3%	32.5%	32.8%	
Mechanical ventilation within the first 24 h after admission*	38.4%	40.6%	38.4%	38.9%	
Unplanned admissions*	22.1%	21.0%	22.1%	21.9%	
Outcome variables					β (95% CI)
Length of stay in hours* (M, SD)	103 (215)	109 (203)	94 (216)	97 (210)	-5.02 (-25.47 to 15.42) [†] OR (95% CI)
Readmissions within 24 h in this period (%) [*]	1.4%	1.1%	1.6%	1.6%	0.92 (0.43 to 1.96) [‡]
ICU deaths in this period (%) [*]	10.9%	10%	12%	10.1%	0.99 (0.70 to 1.04) [§]

Overview of the descriptive results of the patient outcomes during the premeasurement and postmeasurement period for the intervention and control groups, along with the results of the regression analysis of the outcome variables with regard to the effect of the CRM training.

*The descriptive results of the outcome variables have been adjusted for the crude model and APACHE IV severity score.

[†]The linear regression analysis was adjusted for: APACHE IV severity score, use of mechanical ventilation with admission, use of mechanical ventilation within the first 24 h after admission, whether the admission was scheduled or not.

[‡]The logistic regression analysis was adjusted for: use of mechanical ventilation within the first 24 h after admission, whether the admission was scheduled or not.

[§]The logistic regression analysis was adjusted for: APACHE IV severity score, use of mechanical ventilation with admission, use of mechanical ventilation within the first 24 h after admission, whether the admission was scheduled or not.

CRM, crew resource management; ICU, intensive care unit.

Table 4 Error culture questionnaire

Dimension	Intervention M (SD)		Control M (SD)		Interaction component F
	Pre	Post	Pre	Post	
Social orientation	3.35 (0.57)	3.62 (0.57)	3.75 (0.53)	3.77 (0.59)	13.20*
Awareness	2.49 (0.46)	2.60 (0.55)	2.59 (0.46)	2.48 (0.54)	11.11*
Aversion	2.52 (0.37)	2.48 (0.46)	2.39 (0.41)	2.44 (0.40)	3.53
Mastery	3.56 (0.47)	3.71 (0.44)	3.74 (0.39)	3.74 (0.44)	6.29**

Bold indicates significant results.

Overview of the descriptive (M, SD) results for the Error Culture Questionnaire dimensions, along with the outcome of the interaction component in the repeated measures ANOVA.

* $p < 0.01$; ** $p < 0.05$.

ANOVA, analysis of variance.

Table 5 Overview of the descriptive (M, SD, α) results for job satisfaction and affective commitment, along with the outcome of the interaction component in the repeated measures ANOVA

Dimension	k	Cronbach's α (a)		Intervention M (SD)		Control M (SD)		Interaction component F-value
		Pre	Post	Pre	Post	Pre	Post	
Job satisfaction								
Job itself	4	0.72	0.83	3.55 (0.52)	3.72 (0.58)	3.73 (0.46)	3.72 (0.49)	4.72**
The ICU (design and structure)	4	0.81	0.75	3.10 (0.62)	3.42 (0.53)	3.34 (0.54)	3.31 (0.45)	20.17*
Organisational processes	4	0.76	0.75	3.53 (0.56)	3.69 (0.52)	3.56 (0.48)	3.49 (0.50)	9.76*
In general	1	n/a	n/a	3.82 (0.60)	3.98 (0.64)	4.00 (0.62)	4.03 (0.53)	1.48
Affective commitment to the ICU	6	0.83	0.81	3.10 (0.76)	3.34 (0.68)	3.33 (0.56)	3.40 (0.59)	5.23**

k: number of items within the dimension; F-value: the F-value of the interaction between measurement (pre vs post) and experimental group (intervention vs control).

* $p < 0.01$; ** $p < 0.05$.

ANOVA, analysis of variance; ICU, intensive care unit.

Figure 1 Overview of the results on the four levels of the Kirkpatrick evaluation framework. CRM, crew resource management.

