

Does classroom-based Crew Resource Management training improve patient safety culture? A systematic review

SAGE Open Medicine
2: 2050312114529561
© The Author(s) 2014
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2050312114529561
smo.sagepub.com


Inge Verbeek-van Noord¹, Martine C de Bruijne¹, Nicolien C Zwijnenberg², Elise P Jansma³, Cathy van Dyck⁴ and Cordula Wagner^{1,2}

Abstract

Aim: To evaluate the evidence of the effectiveness of classroom-based Crew Resource Management training on safety culture by a systematic review of literature.

Methods: Studies were identified in PubMed, Cochrane Library, PsycINFO, and Educational Resources Information Center up to 19 December 2012. The Methods Guide for Comparative Effectiveness Reviews was used to assess the risk of bias in the individual studies.

Results: In total, 22 manuscripts were included for review. Training settings, study designs, and evaluation methods varied widely. Most studies reporting only a selection of culture dimensions found mainly positive results, whereas studies reporting all safety culture dimensions of the particular survey found mixed results. On average, studies were at moderate risk of bias.

Conclusion: Evidence of the effectiveness of Crew Resource Management training in health care on safety culture is scarce and the validity of most studies is limited. The results underline the necessity of more valid study designs, preferably using triangulation methods.

Keywords

Patient safety (MeSH), safety culture, Crew Resource Management training, systematic review, hospitals (MeSH)

Date received: 23 December 2013; accepted: 5 March 2014

Background

While health-care workers are educated in settings differing in expertise, educational level, and overarching perspective, in practice they have to work together and are expected to be good team players. Until a decade ago, health-care workers received hardly any training in the area of working in teams and corresponding non-technical skills, while the literature shows that many contributing factors to adverse events are related to miscommunication, a lack of communication and teamwork, and other non-technical skills.¹

The importance of non-technical skills was recognised four decades ago in aviation. As a result, specialised training programmes, like Crew Resource Management (CRM), aimed at minimising the effects of human error by improving non-technical skills, were developed to improve safety-critical behaviours on the flight deck.² CRM typically includes educating teams about the limitations of human performance.³ Operational concepts include inquiry, seeking relevant operational information, assessing personal and peer behaviour, communicating proposed actions, conflict resolution, and decision-making.³⁻⁵

In common with others,^{6,7} Salas et al.⁸ reported that CRM training in aviation resulted in positive reactions, enhanced learning, and desired behavioural change in the cockpit. Due to its face validity, the Institute of Medicine advocated the adoption of CRM to safety and error management in health

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

²NIVEL, Netherlands Institute for Health Services Research, Utrecht, The Netherlands

³Medical Library, VU University Amsterdam, Amsterdam, The Netherlands

⁴Department of Organization Sciences, VU University Amsterdam, Amsterdam, The Netherlands

Corresponding author:

Inge Verbeek-van Noord, Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands.

Email: i.vannoord@vumc.nl

care for creating the necessary safety culture.⁹ Consequently, international health authorities placed a high priority on CRM training as a method to improve patient safety, especially in high-risk areas such as emergency departments, intensive care units, and surgery.^{9,10} As a result, efforts have started to be made to implement CRM in the health-care sector.^{6,8} Evaluations of these programmes generally focus on one or more of the four levels of Kirkpatrick and Kirkpatrick's¹¹ framework for evaluating educational interventions: reactions, learning/knowledge, behaviour, and organisational impact.

Several reviews on medical team training exist.^{12–15} The current review focuses on organisational impact – more specifically the patient safety culture – since the ultimate goal of CRM training is to alter safety culture. It is suggested that a positive, proactive safety culture will lead to fewer adverse events and less patient harm.¹⁶ We systematically reviewed the effects of CRM training on patient safety culture to investigate the effect of CRM training on safety culture, focussing on classroom-based training courses given to health-care teams in hospitals. It includes an extensive description of the validity of the included studies.

Methods

This manuscript adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2009 checklist for reporting of systematic reviews.¹⁷

Data sources

Studies were identified by online searches performed up until 19 December 2012 in four electronic databases: PubMed, Cochrane Library, PsycINFO (via EBSCO) and Educational Resources Information Center (ERIC) (via EBSCO). No limits were applied concerning year of publication or language, but only articles in English were included. References of relevant articles and related reviews were checked by hand.

Selection of articles

Using a predefined set of 10 CRM articles, we (M.C.d.B., I.V.N., and E.P.J.) developed a search strategy containing free-text terms only. Indexing terms (controlled terms) did not identify relevant articles, neither did they detect additional articles compared to the free-text terms. In general, indexation of CRM studies was very heterogeneous and lacked standardisation. The following terms were combined with Boolean 'or': 'crew resource management', 'Medical team training', 'Non-technical skills', 'teamwork training', 'team training', 'teamwork performance', 'team performance', 'team resource management', 'Medical team education', 'teamwork education', 'team education', 'team collaboration', 'team behaviour', 'team behavior',

'team skills', 'teamwork skills', 'team decision making', 'team effectiveness', 'team structure', and 'team competencies'. We excluded the term 'crisis resource management' since this term only revealed studies based on simulation techniques.

During the selection procedure, two researchers (I.V.N. and M.C.d.B.) assessed whether references were 'relevant', 'uncertain', or 'irrelevant' according to the inclusion/exclusion criteria based on title and abstract only. Full texts of abstracts judged relevant were requested. Abstracts judged 'uncertain' were re-evaluated. In the case of disagreement regarding a re-evaluated abstract, the full text was requested. Both researchers judged all full texts separately. In the case of disagreement, the full texts were discussed until consensus was reached.

Second, we performed a hand search on existing systematic reviews about team training and checked for relevant references in the included studies. In a final stage, we selected those studies that used safety culture as an outcome (Figure 1).

Inclusion and exclusion criteria

Studies were eligible for inclusion when the training focussed on health-care teams in hospitals and covered at least two CRM topics (e.g. communication and leadership). We excluded studies evaluating CRM in pre-clinical medical education, outside health care, in primary care, and dental care. CRM training courses based (partly) on high-tech simulation techniques were also excluded, as they are fundamentally different from classroom-based training courses. When evaluation studies compared classroom-based CRM training versus simulation-based CRM training or no training, we neglected results from simulation-based training. Furthermore, we excluded manuscripts based on qualitative research.

Data extraction

Studies with a positive effect of CRM training on safety culture were defined as those studies with statistically significant changes from baseline and/or a control group or changes in safety culture dimensions of 10% or more.¹⁸ Descriptive data (information about the type of training, participants, measurement instruments, follow-up, analyses, and implementation and sustainment strategies) were extracted by I.V.N., while data were checked by N.Z. to confirm whether this had been extracted correctly from manuscripts.

Quality appraisal

The Methods Guide for Comparative Effectiveness Reviews of the Agency for Healthcare Research and Quality¹⁹ was used to assess the risk of bias in the individual studies in the systematic review. The taxonomy of the five core biases of

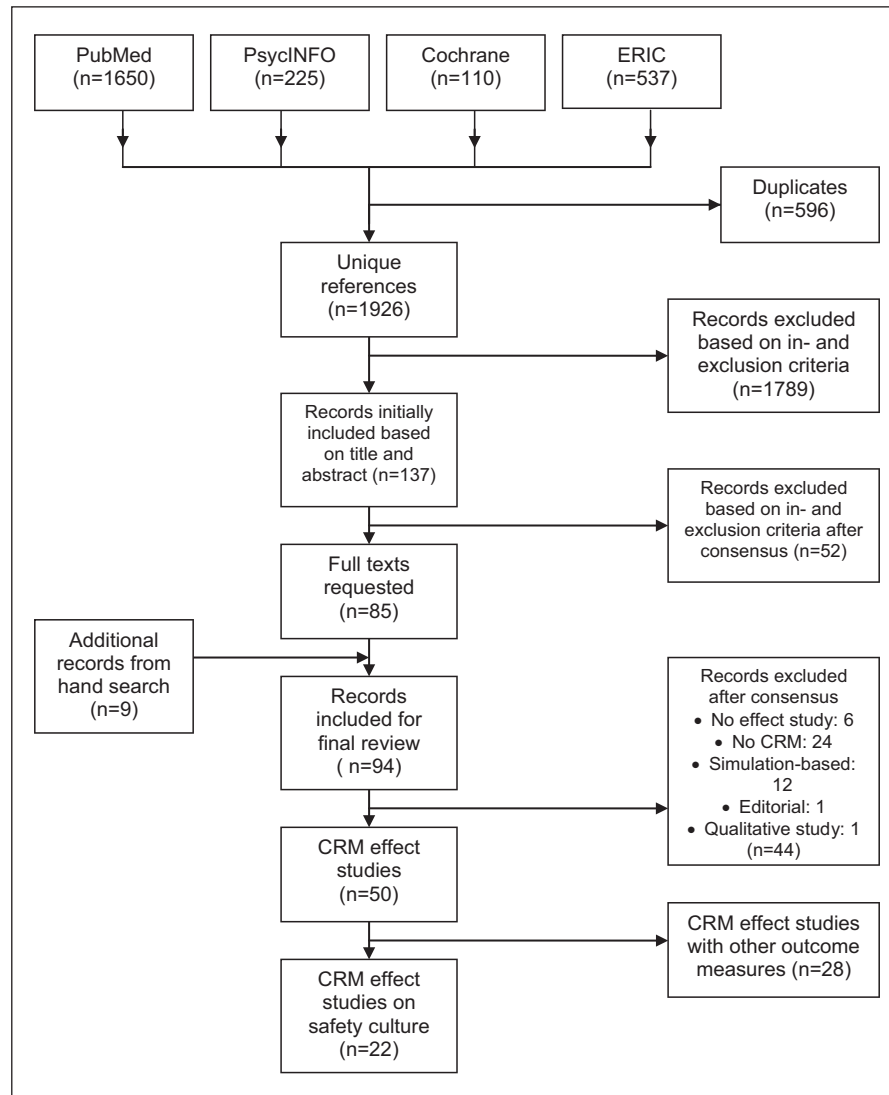


Figure 1. Flowchart of included articles.

ERIC: Educational Resources Information Center; CRM: Crew Resource Management.

the Cochrane Handbook was used, namely, selection bias (including randomisation and blinding bias for randomised controlled trials (RCTs)), performance bias, attrition bias, detection bias, and reporting bias. For RCTs, and cohort- and cross-sectional studies, we used specific criteria according to the description in Table 4 of the Methods Guide for Comparative Effectiveness Reviews (Table 1). Every separate criterion is reported as well as a percentage score of the risk of bias.

Zero points were assigned to low risk of bias, one point to moderate risk of bias, and two points to high risk of bias. The sum of points was divided by the total points possible for all criteria together, multiplied by 100 ('unclears' were disregarded). As a result, 0%–33.2% resembles low risk of bias, 33.3%–66.6% indicates a moderate risk of bias, and 66.7%–100% reflects a high risk of bias.

Results

We retrieved 1650 manuscripts from PubMed, 225 from PsycINFO, 110 from the Cochrane Database, and 537 from the ERIC database, resulting in 1926 unique references. The selection procedure resulted in 50 manuscripts evaluating one of the four levels of Kirkpatrick and Kirkpatrick's¹¹ framework. Of these, 22 reported data on safety culture (Figure 1).

Study and training characteristics

All studies were published from 2006 onwards, with the majority being conducted in the United States. Half of the studies focussed on the operating theatre setting. Of the studies, 16 had uncontrolled designs, of which 10 were single-centre studies. And 10 studies were multicentre studies, 3 of

Table 1. Definitions of the types of biases used for the risk of bias assessment, adjusted for training interventions.

Type of bias	Definition
Selection bias	Systematic differences between baseline characteristics of the groups that arise from self-selection for the intervention, investigator/hospital management-directed selection of intervention, or association of intervention assignments with demographic, clinical, or social characteristics.
Performance bias	Systematic differences in the intervention provided to participants and protocol deviation. Examples include contamination of the control group with the exposure or intervention, unbalanced provision of additional interventions or co-interventions, difference in co-interventions, and the inadequate blinding of providers and participants.
Attrition bias	Systematic differences in the loss of participants from the study and how they were accounted for in the results (e.g. incomplete follow-up, differential attrition). Those who drop out of the study or who are lost to follow-up may be systematically different from those who remain in the study. Attrition bias can potentially change the collective (group) characteristics of the relevant groups and their observed outcomes in ways that affect study results by confounding and spurious associations.
Detection bias	Systematic differences in outcome assessment among groups being compared, including systematic misclassification of the exposure or intervention, covariates, or outcomes because of variable definitions and timings, recall from memory, inadequate assessor blinding, and faulty measurement techniques. Erroneous statistical analysis might also affect the validity of effect estimates.
Reporting bias	Systematic differences between reported and unreported findings (e.g. incomplete reporting of study findings, potential for bias in reporting through source of funding).

which had a controlled design. In total, there were six controlled designs, three in which a control site was used, two in which trained versus non-trained personnel were compared, and one in which the last trained cohort was compared to the first trained cohort before they received training. The TeamSTEPPS training curriculum was implemented on eight occasions. Follow-up measurement varied between 3 months and 4 years. Response percentages differed widely among the studies (range: 19%–96%), as well as the number of trained individuals (range 29–32,150). Implementation and sustainment strategies consisted of, among others, change or leadership teams (usually formed with champion figures), briefings/debriefings, coaching, comprehensive implementation plans, embedding of training within patient safety programmes, structural training of incoming employees, and train-the-trainer modules (Table 2).

Training effects

Table 3 describes the effects of the different studies and their risk of bias. Predominantly, the Hospital Survey on Patient Safety Culture (HSOPSC, 8 times) or the Safety Attitude Questionnaire (SAQ, 11 times) was used to assess the effect of team training on patient safety culture. The results per questionnaire are given below.

Safety Attitude Questionnaire. Regarding studies that used the SAQ as an evaluation method, 25- and 100-point scale scores, item-level differences, and differences in positive responses were reported. Four studies reported results on all SAQ dimensions, two of them finding mainly positive results^{34,38} and two of them finding mainly negative results.^{35,40} Four studies reported only teamwork climate and found positive increases. Safety climate was reported

separately in one study, and a significant positive change was reported.²³ The study of Haller et al.²⁷ reported teamwork climate, safety climate, and stress recognition and found some improvements at item level. In sum, teamwork climate increased in six of the nine studies reporting this dimension. Safety climate changed in a positive direction in four of seven studies that reported that outcome.

Hospital Survey on Patient Safety Culture. As with the SAQ, not all dimensions were evaluated in all studies. Four studies reporting all dimensions found mixed results. Blegen et al.²² found positive changes for all dimensions except for frequency of event reporting (staffing was not reported). Marshall and Manus,¹⁸ in contrast, only found significant changes for four dimensions. Stead et al.³⁶ found statistically significant improvement on two dimensions, but when the cut-off of more than 10% change in positive responses was used, all but one (hospital management support for patient safety) improved. Thomas and Galla³⁷ found positive results in general, although different results were partly seen for the pilot hospital compared to the system-wide evaluation. Four studies reported only selected dimensions of the HSOPSC,^{20,26,30,39} and except the study of Weaver et al.,³⁹ all found positive changes in those selected dimensions. Gore et al.²⁶ did not find positive changes when outcomes were regarded as continuous as opposed to dichotomous (positive answers); when regarded as dichotomous, all selected dimensions changed.

Other questionnaires. Other questionnaires used to assess the effect of team training on patient safety culture were a Brief Teamwork Perception Questionnaire, and the Team Assessment Questionnaire, in which we only considered the domains ‘Leadership’ and ‘Climate’ as relevant for patient safety culture. Castner et al.²⁵ found a result on Leadership.

Table 2. Characteristics of the included studies.

Source	Country	Name of training	Site (n trainees)/ professions	Study design (n centres)	Evaluation instrument	Response rates (%)	Follow-up (months)	Implementation/sustainment strategy
Armour et al. ²⁰	USA	TeamSTEPS	OR (NS)/whole OR team	Single-centre, uncontrolled before–after	HSOPSC, TAQ	NS	9	Leadership team, special training for programme coaches
Bleakley et al. ²¹	UK	NS	OR (about 150)/ whole OR team except anaesthetists working on both sites	Single-centre, controlled before–after	SAQ	Before: 73, after: 68	12	Champion team training, briefing–debriefing, close-call reporting
Blegen et al. ²²	USA	TOPS training	Inpatient unit (454)/ nurses, physicians, pharmacists, therapists, administrators, directors, managers and others	Multicentre (3), uncontrolled before–after	HSOPSC	Before: 96, after: 81	9–12	Creation of Unit Safety Teams consisting of champions and invited or volunteered providers
Carney et al. ²³	USA	VHA, MTT	OR (NS)/all OR personnel	Multicentre (101), uncontrolled before–after	SAQ	Before: 74, after: 36	9–11	Implementation team, briefings and debriefings
Carney et al. ²⁴	USA	VHA, MTT	OR (NS)/OR nurses and OR physicians	Multicentre (101), uncontrolled before–after	SAQ	Before: 74, after: 36	9–11	Implementation team, briefings and debriefings
Castner et al. ²⁵	USA	TeamSTEPS	Entire hospital (1204)/ nurses	Multicentre (5), cross-sectional controlled design (trained vs non-trained)	Brief T-TPQ	Trained: 19, non-trained: 27	NA	Training was mandatory for new employee orientation, master trainers (train-the-trainer principle)
Gore et al. ²⁶	USA	Commercial training	OR (NS)/all OR personnel	Single-centre, uncontrolled before–after	HSOPSC	Before: 35, after: 28	8	Preoperative briefing was developed
Haller et al. ²⁷	CH	Commercial training	L&D (239)/nurses, midwives, physicians, technicians, managers	Single-centre, cross-sectional controlled design (3rd and 2nd vs 1st period)	SAQ-L&D	95	NA	Team improvement strategies were defined
Halverson et al. ²⁸	USA	NS	OR (1150)/all OR personnel	Single-centre, uncontrolled before–after	PTS	NS	6	Train-the-trainer, intraoperative coaching to facilitate briefing and debriefing
Mahoney et al. ²⁹	USA	TeamSTEPS	Mental health (284)/ physicians, nurses, psychologists, administrators	Single-centre, uncontrolled before–after	TAQ	Before: 36, after: 47	Ca 12	Kick-off team, implementation plan, train-the-trainer principle, luncheon debriefings, TeamSTEPS tips for the day posted on Intranet, Trainer and Champion awards, orientation training for new hires, TeamSTEPS on Quality Council monthly agenda
Marshall and Manus ¹⁸	USA	Commercial training	OR (688)/all OR personnel	Multicentre (5), uncontrolled before–after	HSOPSC	Before: 83, after: 71	8–16	OR team briefing models were developed, observation and coaching of OR teams

(Continued)

Table 2. (Continued)

Source	Country	Name of training	Site (n trainees)/professions	Study design (n centres)	Evaluation instrument	Response rates (%)	Follow-up (months)	Implementation/sustainment strategy
Mayer et al. ³⁰	USA	TeamSTEPS	PICU (85), SICU (84), respiratory therapy (90)/all staff	Single-centre, uncontrolled before, mid, and after	HSOPSC	PICU – before: 21, mid: 39, and after: 50; SICU – before: 22, mid: 43, and after: 44	NS	Change team that completed Master Training (train-the-trainer). Members of change team served as unit-based coaches after training. Change team served as steering group for implementation and trainers of frontline staff.
McCulloch et al. ³¹	UK	NS	OR (54)/surgeons, anaesthetists, nurses	Single-centre, uncontrolled before–after	SAQ-OR	NS	3	On-site coaching by aviation CRM trainers twice-weekly for 3 months
Meliones et al. ³²	USA	Health System's TTP	PICU (NS)/all PICU professions	Single-centre, uncontrolled before–after	SAQ	NS	6	Embedded in comprehensive patient safety programme, improvement initiatives derived from training were implemented
Pettiker et al. ³³	USA	Aviation CRM	L&D (289)/physicians, nurses, administrators, assistants	Multicentre (n = 4), uncontrolled before–after	SAQ	Before: 89, between: 95, 24 months: 94, 48 months: 72	NS, ca 12, 24, and 48	Embedded in comprehensive patient safety programme, new employees hired received training shortly after starting work
Pratt et al. ³⁴	USA	NS	L&D (220)/L&D staff	Single-centre, cross-sectional controlled design (trained vs not trained)	SAQ-L&D	NS	Ca 48	Implementation plan, steering team took steps to sustain behavioural process (emails/staff meetings to spread news, praising clinicians, refresher training), new staff required to attend training
Riley et al. ³⁵	USA	TeamSTEPS	L&D (60)/L&D staff	Multicentre (3), before–after controlled	SAQ	NS	12	Unclear
Stead et al. ³⁶	AU	TeamSTEPS	Mental health service (60)/mental health staff	Single-centre, uncontrolled before–after	HSOPSC	Before: 75, after: 76	NS, max 5	Train-the-trainer principle, development of change team
Thomas and Galla ³⁷	USA	TeamSTEPS	Acute care facility (1300), whole system (32,150)/all types of professions	Multicentre (n = 15), uncontrolled before–after	HSOPSC	NS	2 and 3 years	Align training with goals, train-the-trainer programme, comprehensive implementation plan and integration plan (e.g. briefings and leadership participation), Change Team implementation team, briefings/debriefings, consultative interviews
Watts et al. ³⁸	USA	VA, NCPS, MTT	OR (NS)/all OR staff	Multicentre (63), uncontrolled before–after	SAQ	Before: 76, after: 50	8	

Table 2. (Continued)

Source	Country	Name of training	Site (n trainees)/professions	Study design (n centres)	Evaluation instrument	Response rates (%)	Follow-up (months)	Implementation/sustainment strategy
Weaver et al. ³⁹	USA	TeamSTEPS	OR (29)/whole OR teams	Multicentre (n = 2), before–after controlled	HSOPSC	Only numbers are known	I	Train-the-trainer programme, briefings, orientation training to new employees, steering group, integration into curriculum
Wolf et al. ⁴⁰	USA	VHA, MTT	OR (NS)/whole OR team	Single-centre, uncontrolled before–after	SAQ	Only numbers are known	12–17	Formation of implementation team, briefings/debriefings protocol

NS: not specified; VHA: Veterans' Health Administration; VA: Veterans Affairs; NCPS: National Center for Patient Safety; MTT: Medical Team Training; USA: United States of America; UK: United Kingdom; CH: Switzerland; AU: Australia; OR: Operating Room; L&D: Labour and Delivery; PICU: Paediatric Intensive Care Unit; HSOPSC: Hospital Survey on Patient Safety Culture; SAQ: Safety Attitude Questionnaire; T-TPQ: TeamSTEPS Teamwork Perception Questionnaire; PTS: Perception of Teamwork Survey; TAQ: Team Assessment Questionnaire; TTP: Team Training Programme; CRM: Crew Resource Management.

Mahoney et al.²⁹ found a positive improvement on Climate, where Armour et al.²⁰ did not. Halverson et al.²⁸ reported results at item level and observed that 14 of 19 items improved.

Quality appraisal

The studies showed an average risk of bias percentage of 43.7% (standard deviation (SD) 15.3%), indicating that on average the studies had a moderate risk of bias. Seven studies had a low risk of bias (<33.3%),^{21,22,25,27,31,38,39} but three of them did not describe all the bias criteria^{31,38,39} (Table 3); 14 studies had a moderate risk of bias (between 33.3% and 66.7%),^{18,20,23,24,26,28–30,33–37,40} 11 of which^{18,20,23,24,28,30,34–37,40} did not describe all bias criteria (see Appendix 1 for criteria). One study had a high risk of bias (≥66.7%).³² At first sight, it seemed that controlled studies had less risk of bias in general, but a non-parametric independent-sample test did not demonstrate these differences as being significant (p = 0.19)

Selection bias comprised items about allocation and the analyses and design regarding modifying and confounding variables. Nine studies^{26,28,29,32–34,36,37,40} had a high risk of selection bias, mainly because the design and analyses did not take into account possible confounding and modifying variables. Fisher's exact test showed that single-centre uncontrolled studies more often had a high risk of selection bias compared to other study designs (p = 0.494). Performance risk of bias was high in two studies^{18,20} and considered low in three.^{21,31,39} Attrition bias concerned the loss of follow-up of respondents. If attrition is a concern, missing data have to be handled appropriately according to the risk of bias assessment format. A high risk of attrition bias was found on eight occasions,^{23–26,29,30,39,40} and in seven studies, it was unclear how missing data were handled and/or what the response rates were.^{20,28,31,32,34,35,37} The risk of detection bias was considered high in five studies,^{18,32,36,37,40} while it was considered low in seven.^{21,22,25,26,35,38,39} Fisher's exact test revealed that multicentre controlled design more often had a low risk of detection bias compared to the other study designs (p = 0.225). Reporting bias concerned predefined and reported outcome variables. In two studies,^{32,35} outcome variables were prespecified but not all reported, which gave them a high risk of reporting bias classification.

Discussion

This systematic review explored the effects of a classroom-based CRM training on safety culture. In total, 22 studies were included whose effect was mainly evaluated by means of the SAQ or the HSOPSC. Uncontrolled studies in our systematic review all found positive effects, although the magnitude of effects varied across the studies. Two controlled studies that used a control group found no training effects. All but one of the cross-sectional controlled studies found some effects. Risk of bias assessment revealed that in

Table 3. Effects of CRM training described in the included studies and the risk of bias assessment of the included studies.^a

Source	Outcome measure – safety culture dimension(s)	Effects	Risk of bias assessment ^b						
			Selection	Performance	Attrition	Detection	Reporting	Risk of bias (%)	Unclear items (N)
Armour et al. ²⁰	Teamwork within unit	53.2 → 62.7 (out of 100)	1	2	–	1	1	50	3
Bleakley et al. ²¹	Communication openness (other dimensions not reported) Teamwork climate (other dimensions not reported)	47.5 → 62.7 (out of 100) Intervention: 15.3 → 16.5 (out of 25) Control: NS	1	0	0	0	0	20	0
Blegen et al. ²²	Teamwork within units Organisational learning Supervisor/manager expectations Hospital management support for safety Communication openness Error feedback and communication Non-punitive response to error Teamwork across units Hospital handoffs and transitions Overall perception of safety (staffing: not reported)	3.83 → 3.95 (out of 5) 3.53 → 3.81 (out of 5) 3.41 → 3.76 (out of 5) 3.51 → 3.81 (out of 5) 3.44 → 3.63 (out of 5) 3.32 → 3.51 (out of 5) 2.86 → 3.15 (out of 5) 3.36 → 3.51 (out of 5) 2.71 → 2.93 (out of 5) 3.02 → 3.29 (out of 5)	0	0	0	0	0	25	0
Carney et al. ²³	Safety climate (other dimensions not reported)	High complexity: 7 out of 7 items improved Medium complexity: 7 out of 7 items improved	1	1	2	1	1	44	1
Carney et al. ²⁴	Teamwork climate (other dimensions not reported)	Nurses: improvement in 5 out of 6 items Physicians: improvement in 6 out of 6 items No overall teamwork climate score reported	1	1	2	1	1	44	1
Castner et al. ²⁵	Leadership (no other relevant dimensions in questionnaire)	Higher leadership scores in trained group	0	1	2	0	0	33	0
Gore et al. ²⁶	Teamwork Error reporting	Continuous outcome Positive response improved 4/4 items improved 10/11 items improved (!)	2	1	2	0	1	50	0

Table 3. (Continued)

Source	Outcome measure – safety culture dimension(s)	Effects	Risk of bias assessment ^b					Risk of Unclear bias (%)	Unclear items (N)
			Selection	Performance	Attrition	Detection	Reporting		
Haller et al. ²⁷	Safety climate (other dimensions not reported)	2/11 items improved (!) 8/13 items improved (!)	0	1	0	1	1	22	0
	Teamwork climate	Improvement in 2/12 questions							
	Safety climate	Improvement in 1/12 questions							
	Stress recognition (other dimensions not reported)	Improvement in 3/8 questions							
Halverson et al. ²⁸	Perception of teamwork dimensions in questionnaire	14/19 items improved	2	1	–	1	0	50	1
	Climate and atmosphere dimensions in questionnaire	3.68 → 3.97 (out of 5)	2	1	2	1	0	50	0
Mahoney et al. ²⁹	Teamwork within units	Average change in positive responses (%)	1	2	0	2	0	56	2
Marshall and Manus ¹⁸	Hospital handoffs and transitions	12 (range 0–16)							
	Frequency of events reported	10 (range 2–19)							
	Staffing	11 (range 1–19)							
	Overall perceptions of safety	12 (range 0–20)							
Mayer et al. ³⁰	Communication openness	PICU, SICU, and whole group: improvement	0	1	2	1	0	35	2
	Teamwork within unit (other dimensions not reported)	PICU, SICU, and whole group: improvement							
McCulloch et al. ³¹	Teamwork climate (other dimensions not reported)	64.1 → 69.2 (out of 100)	1	0	–	1	0	21	1
	Teamwork climate (other dimensions not reported)	67.3% to 86.9% positive responses	2	1	–	2	2	75	2
Pettker et al. ³³	Teamwork climate	39% to 63% positive responses	2	1	0	1	0	44	0
	Safety climate	33% to 63% positive responses							
Pratt et al. ³⁴	Job satisfaction	39% to 53% positive responses							
	Perception of management	10% to 37% positive responses							
	Safety climate (other dimensions not reported)	Percentage of positive responders is higher for L&D staff than entire hospital	2	1	–	1	1	57	1
Riley et al. ³⁵	All SAQ dimensions	NS	1	1	–	0	2	50	4
Stead et al. ³⁶	Frequency of event reporting	28% to 53% positive responses	2	1	0	2	0	50	1

(Continued)

Table 3. (Continued)

Source	Outcome measure – safety culture dimension(s)	Effects	Risk of bias assessment ^b							
			Selection	Performance	Attrition	Detection	Reporting	Risk of Unclear bias (%) items (N)		
Thomas and Galla ³⁷	Organisational learning – continuous improvement	49% to 79% positive responses	2	1	–	2	1	64	1	
	Increase in percentage of positive responses after 3 years									
	Hospital handoffs and transitions	11.3								
	Hospital management support for patient safety	11								
	Non-punitive response to error	15.9								
	Organisational learning – continuous improvement	11.7								
	Overall perceptions of safety	11.8								
	Staffing	15.8								
	Supervisor/manager expectations and actions promoting patient safety	10.9								
	Teamwork across units	14.1								
	Teamwork within units	11.9								
	System-wide results		Improvement							
	Feedback and communication about error		Improvement							
	Frequency of events reported		Improvement							
Watts et al. ³⁸	Hospital handoffs and transitions									
	Staffing									
	Teamwork across units									
	Organisational learning									
	Teamwork within units									
	Teamwork climate		0	1	1	0	0	25	1	
	Safety climate									
	Job satisfaction		65.8 → 72.1 (out of 100)							
	Stress recognition		67.4 → 72.9 (out of 100)							
	Perception of management		72.1 → 73.5 (out of 100)							
	Work conditions		68.2 → 69.7 (out of 100)							
	Teamwork within units		56.1 → 63.7 (out of 100)							
			60.1 → 64.3 (out of 100)							
			Positive increase in percentage of positive responses in intervention and control group	1	0	2	0	0	33	1
Weaver et al. ³⁹	Communication openness	No change								
	Feedback and communication about error	No change								
	Overall perceptions of safety (other dimensions not reported)	No change								

Table 3. (Continued)

Source	Outcome measure – safety culture dimension(s)	Effects	Risk of bias assessment ^b						
			Selection	Performance	Attrition	Detection	Reporting	Risk of bias (%)	Unclear items (N)
Wolf et al. ⁴⁰	Perceptions of management Working conditions	Positive increase in percentage of positive responses Positive increase in percentage of positive responses	2	1	2	2	0	64	1

NS: non-significant; L&D: Labour and Delivery; PICU: Paediatric Intensive Care Unit; SAQ: Safety Attitude Questionnaire; SICU: Surgical Intensive Care Unit; CRM: Crew Resource Management.
^aOnly statistically significant or relevant (more than 10% change) effects are reported. If not all dimensions of the particular questionnaires are reported, this is mentioned, otherwise no effects were found.

^b0: low risk of bias; 1: moderate risk of bias; 2: high risk of bias; (-): unclear. Percentages are calculated by assigning zero points to low, one point to moderate, and two points to high risk of bias per criterion (criteria not shown separately in this table). The sum of points is divided by the total possible points for all criteria together times 100 ('unclears' were disregarded).

(!) Note that number of items changed when the outcomes were regarded as dichotomous. We could not discover which numbers are right.

general, studies were at moderate risk of bias, with selection bias and attrition bias being the most common biases. Results also showed that single-centre uncontrolled designs were at higher risk of selection bias than other designs and that multicentre studies had a lower risk of detection bias than other designs.

There are several possible reasons why the results of the controlled studies were different from those of the uncontrolled studies. First, single-centre studies usually involve hospital-driven initiatives. In essence, these differ from externally evaluated training programmes since internally driven improvement projects are likely to have more support from upper and middle managers and staff. According to Salas et al.,⁴¹ organisational support is one of the critical success factors determining the success of a training course. Additionally, uncontrolled studies allow more variance in content and implementation strategies than controlled studies. Embedment in organisational goals and aims will motivate frontline care leaders and managers to commit to CRM training principles.⁴¹ Second, compared to controlled designs, training effects in uncontrolled designs may have been more confounded by context factors (e.g. staffing profiles, departmental activities concerning patient safety) or by the fact that patient safety is on the national research agenda in many countries. Context may influence the interpretation of results due to possible beta statistical (type II) errors since outcomes could have been influenced by factors other than the intervention alone. Nevertheless, our risk of bias assessment did not reveal that this aspect played a role. Additionally, context may influence the fidelity of the implementation of the trained principles in practice.⁴² Third, the timing of the measurement may be of influence on the mixed results of the studies, although no clear pattern can be discovered between the magnitude of effects found in studies in our systematic review and their follow-up periods.

We noted some striking findings in this review. First, none of the multicentre or multisite studies made corrections for the clustering of responses within units. Effects may therefore be overestimated since clustering of responses decreases power when intraclass correlation coefficients (ICCs) are high. Concerning the clustering of responses within units or hospitals, Smits et al.⁴³ reported ICCs for the HSOPSC that ranged from 4.3 to 31.7 for the unit level and 0.0 to 6.2 for the hospital level, with 15 as a threshold for high clustering.⁴⁴ Second, the relatively high frequency of high risk of attrition bias suggests that the biggest challenge lies within the prevention of losses to follow-up of respondents or in achieving a high response in the case of cross-sectional studies. Third, the outcome data were often handled as a dichotomous outcome, and consequently, results moved sooner towards statistical significance, as shown in the study by Gore et al.²⁶ Furthermore, when the cut-off of a 10% change in positive responses is used, there is a chance that results are regarded as significant, while statistical tests will

not show any significant results, as shown in the study by Stead et al.³⁶

In general, results of the studies varied widely, and keeping in mind the possible bias, we are cautious in drawing firm conclusions. The possibility of publication bias supports this feeling. When an intervention study does not find an effect, study quality evaluations are more stringent, focusing on the appropriateness of the study design, measurements, and methodology. Thus, intervention studies with negative results have a lower chance of publication.⁴⁵ By contrast, single-centre controlled studies have more chance of publication as these studies show larger effects.⁴⁶

Regarding generalisability, we must take into account that safety culture is a subcomponent of organisational culture and will thereby be influenced by the dominant organisational culture. Organisational culture reflects shared behaviours, beliefs, attitudes, and values regarding goals, functions, and procedures.⁴⁷ As with organisational culture, beliefs, values, and attitudes towards safety culture could vary between individuals. The same variance is possible between departments and within disciplines, which hampers the generalisability of results to other departments and types of departments, as those settings will have profound differences in initial safety culture as well as contextual features. The variety of evaluation methods and designs used to track changes in culture elements aggravates this problem.

In addition to this, the variety in CRM training concepts and the manner in which they are trained makes it hard to pinpoint which training elements are related to specific culture changes. CRM, in general, is focussed on non-technical skills, but different training underscores different concepts or use different training forms. Moreover, expert trainers might adapt their programme, exercises, and feedback to the knowledge, skills, and dynamics in the group. Extensive descriptions of CRM training interventions are therefore a prerequisite when more in-depth analyses after correlations or causations are desirable. Another recommendation would be to also provide thorough descriptions of the trained participants, to be able to gain more insight into for which teams CRM training potentially works.

With respect to methodology, using questionnaires to assess safety culture might not always be the best choice. This method is appropriate within the analytical approach, which assumes that culture is something an organisation *has*. Although this analytical approach⁴⁸ is useful for comparative research, the individual responses to the questionnaire still depend on subjective individuals. We believe that academic and pragmatic approaches – which acquire in-depth interviews and observations before drawing inferences – are lacking. These approaches provide the context for the data resulting from the analytical approach.⁴⁸ Using a triangulation methodology to study intervention effects and safety culture in combination would strengthen inferences about

possible intervention effects, as well as the external and internal validity of the safety culture construct.⁴⁷

Internationally, the impact of team training on secondary outcomes such as adverse outcomes and safety culture change has to be evaluated with highly reliable study designs. This is challenging, especially in the health-care environment where clinical practice is influenced by a variety of highly uncontrollable factors.⁴⁹ Non-controlled before–after evaluations of the specified secondary outcomes might seem the most realisable of all study designs. However, as mentioned previously, in these studies, effects could possibly be attributed to other developments within or outside the organisation since it is harder to distinguish between cause and effect.⁵⁰ One might suggest that controlled clustered (randomised) studies would be a suitable solution. A sufficiently large intervention and control group is a prerequisite.⁵⁰ For example, Nielsen et al.⁵¹ have considered that 11–13 sites per group are needed to have 80% power to detect a 40% reduction in adverse outcomes at labour and delivery units. Another appropriate design for the evaluation of CRM effects may be the stepped wedge design in which sites will act as their own control. This will also reduce the risk of bias. Advantages and disadvantages of the stepped wedge design are mainly of a practical nature, that is, its design suits situations in which interventions allow for a phased introduction, but it demands a laborious and extended amount of data collection.⁵⁰

We are aware of the delay in the publication process since December 2012. We performed a quick scan of the literature in 2013 and the beginning of 2014 in MEDLINE, the database in which all the included articles can be found. Probably four studies have been published since December 2012 that we would have included in this review.^{52–55} Three studies were quasi-experimental^{52,53,54} and one study was a randomised controlled trial,⁵⁵ two of them being small.^{53,55} Since these studies did not show extreme findings at first sight, we do not expect that an update of our review will materially change the results.

In sum, we conclude that evidence of the effectiveness of CRM training in health care in terms of improved safety culture is scarce and the validity of most studies is limited, due to the predominant use of uncontrolled study designs. Although it might be easier to comply with critical success factors for team training with single-centre evaluations, the results underline the necessity of a control group to reduce the risk of bias. In addition to that, more in-depth measurements of context and triangulation methods to analyse these in combination with primary outcomes will help to acquire insight into the working mechanisms of the CRM training and the influential role of context.

Acknowledgements

This study was part of the National Patient Safety Program for Hospitals in the Netherlands.

Declaration of conflicting interests

The authors declare that they have no competing interests.

Funding

Financial support for this study was provided by the Dutch Ministry for Health, Welfare and Sports (grant no. 1075879).

References

- Helmreich RL. On error management: lessons from aviation. *BMJ* 2000; 320(7237): 781–785.
- Cooper GE (ed.). *Resource management on the flight deck*. Moffett Field, CA: NASA, 1980.
- Kosnik LK. The new paradigm of crew resource management: just what is needed to re-engage the stalled collaborative movement? *Jt Comm J Qual Improv* 2002; 28(5): 235–241.
- Oriol MD. Crew resource management: applications in health-care organizations. *J Nurs Adm* 2006; 36(9): 402–406.
- Thomas EJ, Sherwood GD and Helmreich RL. Lessons from aviation: teamwork to improve patient safety. *Nurs Econ* 2003; 21(5): 241–243.
- Baker DP, Gustafson S, Beaubien JM, et al. Team training in health care: a review of team training programs and a look toward the future. *Adv Patient Saf* 2008, http://www.air.org/sites/default/files/downloads/report/adv_pub_safety_0.pdf
- Edkins GD. A review of the benefits of aviation human factors training. *Hum Factors Aero Saf* 2002; 2: 201–216.
- Salas E, Wilson KA, Burke CS, et al. Does crew resource management training work? An update, an extension, and some critical needs. *Hum Factors* 2006; 48(2): 392–412.
- Kohn LT, Corrigan JM and Donaldson MS. *To err is human: building a safer health system*. Washington, DC: Institute of Medicine, 1999.
- Pizzi L, Gofarb NI and Nash DB. *Crew resource management and its application in medicine*. Report no. 01E058, 2001. Rockville, MD: Agency for Healthcare Research and Quality.
- Kirkpatrick DL and Kirkpatrick JD. *Evaluating training programs: the four levels*. 3rd ed. San Francisco, CA: Berrett-Koehler, 2006.
- Manser T. Teamwork of patient safety in dynamic domains of healthcare: a review of literature. *Acta Anaesthesiol Scand* 2009; 53: 143–151.
- Rabol LI, Ostergaard D and Mogensen T. Outcomes of classroom-based team training interventions for multiprofessional hospital staff. A systematic review. *Qual Saf Health Care* 2010; 19(6): e27.
- Weaver SJ, Lyons R, DiazGranados D, et al. The anatomy of health care team training and the state of practice: a critical review. *Acad Med* 2010; 85: 1746–1760.
- Zelster MV and Nash DB. Approaching the evidence basis for aviation-derived teamwork training in medicine. *Am J Med Qual* 2010; 25(1): 13–23.
- Committee on Quality of Health Care in America. *Crossing the quality chasm: a new health system for the 21st century*. Washington, DC: Institute of Medicine, 2001.
- Moher D, Liberati A, Tetzlaff J, et al.; The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; 339: b2535.
- Marshall DA and Manus DA. A team training program using human factors to enhance patient safety. *AORN J* 2007; 86(6): 994–1011.
- Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews. March 2012. AHRQ Publication no. 12-EHC047-EF, http://effectivehealthcare.ahrq.gov/ehec/products/322/998/MethodsGuideforCERs_Viswanathan_IndividualStudies.pdf
- Armour FR, Bramble JD and McQuillan R. Team training can improve operating room performance. *Surgery* 2011; 150(4): 771–778.
- Bleakley A, Boyden J, Hobbs A, et al. Improving teamwork climate in operating theatres: the shift from multiprofessionalism to interprofessionalism. *J Interprof Care* 2006; 20(5): 461–470.
- Blegen MA, Sehgal NL, Alldredge BK, et al. Republished paper: improving safety culture on adult medical units through multidisciplinary teamwork and communication interventions: the TOPS Project. *Postgrad Med J* 2010; 86(1022): 729–733.
- Carney BT, West P, Neily J, et al. Changing perceptions of safety climate in the operating room with the Veterans Health Administration medical team training program. *Am J Med Qual* 2011; 26(3): 181–184.
- Carney BT, West P, Neily JB, et al. Improving perceptions of teamwork climate with the Veterans Health Administration medical team training program. *Am J Med Qual* 2011; 26(6): 480–484.
- Castner J, Foltz-Ramos K, Schwartz DG, et al. A leadership challenge: staff nurse perceptions after an organizational TeamSTEPPS initiative. *J Nurs Adm* 2012; 42(10): 467–472.
- Gore DC, Powell JM, Baer JG, et al. Crew resource management improved perception of patient safety in the operating room. *Am J Med Qual* 2010; 25(1): 60–63.
- 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(4): 254–263.
- Halverson AL, Andersson JL, Anderson K, et al. Surgical team training: the Northwestern Memorial Hospital experience. *Arch Surg* 2009; 144(2): 107–112.
- Mahoney JS, Ellis TE, Garland G, et al. Supporting a psychiatric hospital culture of safety. *J Am Psychiatr Nurses Assoc* 2012; 18(5): 299–306.
- Mayer CM, Cluff L, Lin WT, et al. Evaluating efforts to optimize TeamSTEPPS implementation in surgical and pediatric intensive care units. *Jt Comm J Qual Patient Saf* 2011; 37(8): 365–374.
- 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(2): 109–115.
- Meliones JN, Alton M, Mericle J, et al. *10-year experience integrating strategic performance improvement initiatives: can the balanced scorecard, Six Sigma®, and team training all thrive in a single hospital?* Rockville, MD: Agency for Healthcare Research and Quality, 2008.

33. Pettker CM, Thung SF, Raab CA, et al. A comprehensive obstetrics patient safety program improves safety climate and culture. *Am J Obstet Gynecol* 2011; 204(3): 216.
34. Pratt SD, Mann S, Salisbury M, et al. John M. Eisenberg Patient Safety and Quality Awards. Impact of CRM-based training on obstetric outcomes and clinicians' patient safety attitudes. *Jt Comm J Qual Patient Saf* 2007; 33(12): 720–725.
35. Riley W, Davis S, Miller K, et al. Didactic and simulation nontechnical skills team training to improve perinatal patient outcomes in a community hospital. *Jt Comm J Qual Patient Saf* 2011; 37(8): 357–364.
36. Stead K, Kumar S, Schultz TJ, et al. Teams communicating through STEPPS. *Med J Aust* 2009; 190(11 Suppl.): S128–S132.
37. Thomas L and Galla C. Building a culture of safety through team training and engagement. *BMJ Qual Saf* 2013; 22(5): 425–434.
38. Watts BV, Percarpio K, West P, et al. Use of the Safety Attitudes Questionnaire as a measure in patient safety improvement. *J Patient Saf* 2010; 6(4): 206–209.
39. Weaver SJ, Rosen MA, DiazGranados D, et al. Does teamwork improve performance in the operating room? A multilevel evaluation. *Jt Comm J Qual Patient Saf* 2010; 36(3): 133–142.
40. Wolf FA, Way LW and 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(3): 477–483.
41. Salas E, Almeida SA, Salisbury M, et al. What are the critical success factors for team training in health care? *Jt Comm J Qual Patient Saf* 2009; 35(8): 398–405.
42. Brown C, Hofer T, Johal A, et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 3. End points and measurement. *Qual Saf Health Care* 2008; 17(3): 170–177.
43. Smits M, Wagner C, Spreeuwenberg P, et al. Measuring patient safety culture: an assessment of the clustering of responses at unit level and hospital level. *Qual Saf Health Care* 2009; 18: 292–296.
44. Goldstein H. *Multilevel statistical models*. New York: Halstead Press, 1995.
45. Dwan K, Altman DG, Arnaiz JA, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* 2008; 3(8): e3081.
46. Dechartres A, Boutron I, Trinquart L, et al. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. *Ann Intern Med* 2011; 155(1): 39–51.
47. Cooper MD. Toward a model of safety culture. *Safety Sci* 2000; 36: 111–136.
48. Guldenmund FW. (Mis)understanding safety culture and its relationship to safety management. *Risk Anal* 2010; 30(10): 1466–1480.
49. Vincent C, Taylor-Adams S and Stanhope N. Framework for analysing risk and safety in clinical medicine. *BMJ* 1998; 316(7138): 1154–1157.
50. Brown C, Hofer T, Johal A, et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 2. Study design. *Qual Saf Health Care* 2008; 17(3): 163–169.
51. Nielsen PE, Goldman MB, Mann S, et al. Effects of teamwork training on adverse outcomes and process of care in labor and delivery: a randomized controlled trial. *Obstet Gynecol* 2007; 109(1): 48–55.
52. Jones F, Pidila P and Powers C. Creating a culture of safety in the emergency department: the value of teamwork training. *J Nurs Adm* 2013; 43: 194–200.
53. Spiva L, Robertson B, Delk ML, et al. Effectiveness of team training on fall prevention. *J Nurs Care Qual* 2014; 29: 164–173.
54. Jones KJ, Skinner AM, High R, et al. A theory-driven, longitudinal evaluation of the impact of team training on safety culture in 24 hospitals. *BMJ Qual Saf* 2013; 22: 394–404.
55. Clay-Williams R, McIntosh CA, Kerridge R, et al. Classroom and simulation team training: a randomized controlled trial. *Int J Qual Health Care* 2013; 25: 314–321.

Appendix 1. Risk of bias criteria used for the risk of bias assessment (Viswanathan et al.¹⁹)

Type of bias	Criterion	Study design		
		RCT	CCT/cohort study	Cross-sectional
Selection bias	Was the allocation sequence generated adequately?	x		
	Was the allocation of treatment adequately concealed?	x		
	Were participants analysed within the groups they were originally assigned to?	x	x	
	Did the study apply inclusion/exclusion criteria uniformly to all comparison groups?		x	x
	Did the strategy for recruiting participants into the study differ across study groups?		x	
Performance bias	Does the design or analysis control account for important confounding and modifying variables through matching, stratification, multivariable analyses, or other approached?	x	x	x
	Did researcher rule out any impact from a concurrent intervention or an unintended exposure that might bias results?	x	x	x
Attrition bias	Did the study maintain fidelity to the intervention protocol?	x	x	x
	If attrition was a concern, were missing data handled appropriately?	x	x	x
Detection bias	Was the length of follow-up different between the groups or was the time period between the intervention/exposure and outcome the same for cases and controls?	x	x	
	Were outcomes assessors blinded to the intervention or exposure status of participants?	x	x	x
	Were intervention exposure assessed using valid and reliable measures implemented consistently across all study participants?	x	x	x
	Were outcomes assessed using valid and reliable measures implemented consistently across all study participants?	x	x	x
	Were confounding variables assessed using valid and reliable measures implemented consistently across all study participants?		x	x
Reporting bias	Were the potential outcomes prespecified by the researchers? Are all prespecified outcomes reported?	x	x	x