

**A Meta-Analysis of the Effectiveness of Crew Resource Management  
Training in Acute Care Domains**

**Angela O’Dea <sup>1</sup>, Paul O’Connor<sup>1</sup>, & Ivan Keogh<sup>2</sup>**

Department of General Practice<sup>1</sup>; Academic Department of Otolaryngology<sup>2</sup>, National  
University of Ireland, Galway, Ireland.

**Cite as: O’Dea, A., O’Connor, P., & Keogh, I. (2014). A meta-analysis of  
the effectiveness of crew resource management in acute care  
domains. *Postgraduate Medical Journal*, 90(1070), 699-708.**

## **A Meta-Analysis of the Effectiveness of Crew Resource Management Training in Acute Care Domains.**

### **ABSTRACT**

**Background.** The healthcare industry has seen an increase in the adoption of team training, such as crew resource management (CRM), to improve teamwork and coordination within acute care medical teams. A meta-analysis was carried out in order to quantify the effects of CRM training on reactions, learning, behaviour and clinical care outcomes. Biases in the research evidence are identified and recommendations for training development and evaluation are presented. **Method.** PUBMED, EMBASE and PsychInfo were systematically searched for all relevant papers. Peer-reviewed papers published in English between January 1985 and September 2013; which present empirically based studies focusing on interventions to improve team effectiveness in acute health care domains were included. **Results.** A total of 20 CRM-type team training evaluation studies were found to fulfil the *a priori* criteria for inclusion in the meta-analysis. Overall, CRM trained participants responded positively to CRM (mean score 4.25 out of maximum of 5), the training had large effects on participants' knowledge ( $d=1.05$ ), a small effect on attitudes ( $d=0.22$ ) and a large effect on behaviours ( $d=1.25$ ). There was insufficient evidence to support an effect on clinical care outcomes or long-term impacts. **Discussion.** The findings support the premise that CRM training can positively impact teamwork in healthcare and provide estimates of the expected effects of training. However, there is a need for greater precision in outcome assessment, improved standardisation of methods and measures, and more robust research design. Stronger evidence of effectiveness will require multi-level, multi-centre, multi-specialty, and longitudinal studies.

## INTRODUCTION

Human factors and poor teamwork are the major contributors to poor performance in high risk industries.<sup>1</sup> Similarly, in healthcare, breakdowns in communication, poor teamwork, lack of leadership, and poor decision making by individuals and teams have all been shown to be major contributors to poor performance.<sup>2-4</sup> Recognising the impact of poor teamwork on safety and performance, the aviation industry introduced team-oriented training called Crew Resource Management (CRM) for pilots in the late 1970s. The purpose of CRM training is to promote safety and enhance efficiency through optimum use of all available resources: equipment, procedures and people.<sup>5</sup> The focus of CRM training is not on technical skills, but rather cognitive and interpersonal skills such as communication, situational awareness, problem solving, decision making, leadership, assertiveness and teamwork. Training is usually designed to develop generalizable, transportable teamwork competencies that learners can apply across different settings and teams. Instructional methods include: information-based methods (e.g., didactic lecture); demonstration-based methods (e.g., behavioural modelling, videos); and practice-based methods (e.g., simulation, role playing).<sup>1</sup>

There have been a number of literature reviews<sup>6-8</sup> and a meta-analysis<sup>9</sup> exploring the effectiveness of CRM across a range of high-risk industries. Overall, these reviews suggest a positive effect of training on participant reactions to training, attitudes, knowledge and behaviours related to teamwork. However there is a lack of evidence regarding the effect of CRM training on objective comes such as errors, accidents, and mishaps.

Healthcare shares many of the characteristics of other high-risk work domains, particularly in relation to the requirement for high functioning teams, the high cost of error and the goal of

risk reduction. In recent years there has been an upsurge in the application of CRM-type training within healthcare; such training provides a systematic methodology for improving the knowledge, skills, attitudes and behaviours that underlie effective teamwork. Critical elements of the training include: the opportunity for practice, formative feedback, and tools to support transfer of training to the daily care environment.

In recent years, a number of systematic reviews evaluating team-training interventions in healthcare have been published.<sup>10-14</sup> Such reviews provide a synthesis of the evidence to date. Overall, the evidence supports the notion that team training can meaningfully improve participant knowledge, attitudes, teamwork processes, clinical care processes and even clinical care outcomes.

Previous reviews show that evaluation of training effectiveness usually involves uncontrolled pre-post evaluation of non-technical skills related to teamwork such as communication, decision making and situational awareness. Evaluations are usually categorised under one or more levels of training evaluation corresponding to Kirkpatrick's<sup>15</sup> hierarchy of training evaluation which includes reactions, learning, behaviour and organisational level outcomes.

The purpose of this paper is to build on the existing body of evidence by applying meta-analyses techniques to quantify the effectiveness of CRM-type team training in healthcare settings. Use of meta-analysis will improve upon previous literature reviews through systematic application of quantitative procedures in order to determine the aggregate size of the effects of CRM type training on individual and clinical care outcomes.<sup>16</sup> With the proliferation of CRM-type team training in healthcare settings, an empirical summation of the literature evaluating the effectiveness of such training within this domain is warranted.

The objectives of the meta-analysis are primarily to determine the aggregate size of the effect of CRM training in acute care settings at four different levels of evaluation: reactions, learning, behaviour and clinical care outcomes. Additionally this analysis seeks to identify biases in the research evidence in order to improve the quality of future CRM training interventions in healthcare and also the quality of evaluations of those interventions.

## **METHOD**

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement<sup>17</sup> was used to guide the reporting of this meta-analysis.

### **Protocol**

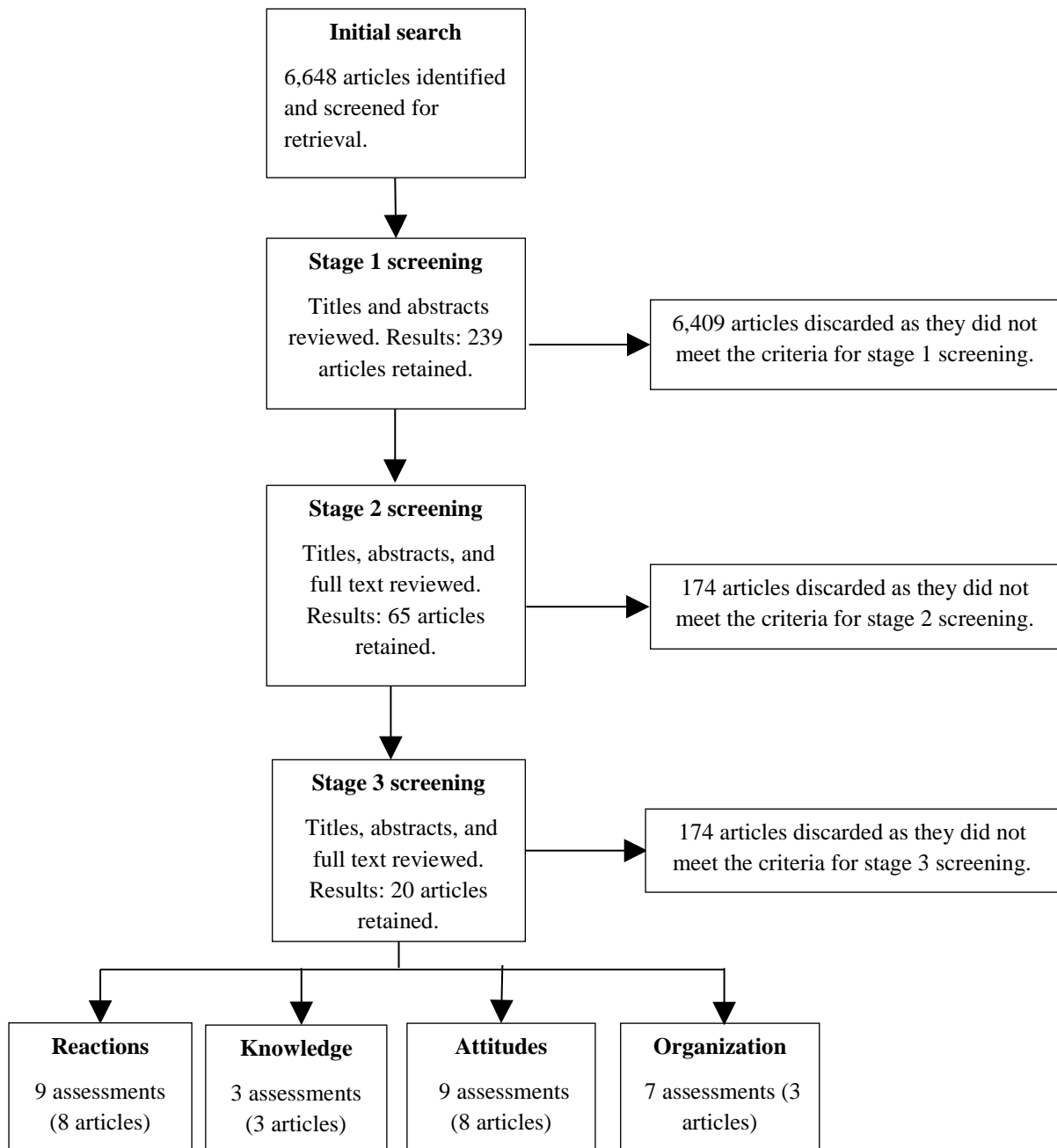
A review protocol and a search strategy following PRISMA guidelines were compiled and revised by the investigators who have expertise in systematic review and meta-analysis methodologies, medical education and clinical care.

### **Eligibility criteria**

All studies included in the meta-analysis met predetermined eligibility criteria: the studies must report CRM-type training interventions that are focused on improving teamwork within healthcare teams in acute care environments;; training effectiveness must be assessed at least one level of Kirkpatrick's evaluation hierarchy<sup>15</sup> (i.e. reactions, learning, behaviour, or organisation).

The exclusion criteria were: training that focuses on specific technical skills or procedures rather than on teamwork. Studies that relate to patient or relative-centred communication or collaboration, studies aimed at administrators, leaders or managers. Figure 1 summarises the process by which the studies were screened for inclusion in the meta-analysis.

Figure 1. PRISMA<sup>17</sup> flow of information through the different phases of the review.



## Identification

The literature search was performed by the authors; it involved a systematic search of PUBMED, Embase and PsychInfo. The search period was from January 1985 until September 2013. The PubMed version of the detailed strategy was used as the basis for the other search strategies, with a targeted hand-search of the references from relevant articles and an auxiliary electronic name search for related articles from individuals and institutions identified as being highly active in the initial search. The PubMed search using the above strategy was: ((patient care team OR nursing team OR institutional management teams OR management quality circles or doctors or nurses or physicians or surgeons or surgical team or anesthesiologist)) AND (((teamwork[Title/Abstract] OR communication[Title/Abstract] OR crisis resource management[Title/Abstract] OR human factors[Title/Abstract] OR crew resource management[Title/Abstract] OR CRM[Title/Abstract] OR team[Title/Abstract] OR nontechnical skill[Title/Abstract] OR behavioral skills[Title/Abstract] OR team dynamics[Title/Abstract] OR TeamSTEPPS[Title/Abstract] OR MED TEAM [Title/Abstract]) AND (training[Title/Abstract])).

## Screening

All of the literature was screened for eligibility by two of the authors (AOD and POC). At each of the stages of screening 10% of the articles were screened by both authors in order to ensure that eligibility criteria were being consistently applied. A total of 100% agreement was found between reviewers in determining eligibility at each stage. The remaining articles at each stage were screened independently. For any articles for which a reviewer was

undecided about eligibility for inclusion, these articles were screened by both authors, and a decision was reached by consensus.

*Initial screening.* A total of 6,648 articles were identified for screening. Articles were screened with reference to the following parameters: population (i.e. learners), setting, programme, outcomes, objective and design. These parameters were used in order to select only training interventions that were focused on improving teamwork within healthcare teams in environments with an acute care focus. Of the initial 6,648 articles, 6409 were excluded and 239 were retained for further screening. The reasons for exclusion are shown in Figure 1.

*Review of design and methods.* The full texts of the remaining 239 articles were retrieved and screened. Articles were reviewed with particular focus on study design and evaluation methods (see Figure 1). Only uncontrolled pre-post design intervention studies were retained for further analysis as this design is by far the most commonly used in the field of research, accounting for approximately 80% of eligible research. Only evaluations that could be categorised as an evaluation of reactions, learning, behaviour, or patient care outcomes were retained.

*Effect size calculation.* Studies were then reviewed in order to determine if sufficient data were available in order to allow effect sizes to be calculated. In total 45 articles with insufficient data to allow for the calculation of an effect size were discarded. A total of 20 articles were retained (see Figure 1).



Finally, Kirkpatrick's evaluation hierarchy<sup>15</sup> was used to group studies based upon the level of evaluation: (Level 1: reactions, Level 2: learning, level 3: behaviour, level 4: clinical care outcomes). The specific inclusion criteria for each level were:

- *Reactions*: An evaluation was coded a measure of reactions if it pertained to likeability, usefulness, or relevance of the training. Responses must be measured on a Likert scale and allow a mean score to be calculated.
- *Learning*: An evaluation was coded as an assessment of learning if knowledge about teamwork or attitudes towards teamwork were assessed. The study must utilise a pre-post design, and must contain sufficient information to allow Cohen's *d* and the standard error (SE) to be calculated.
- *Behaviour*: For a study to be coded as an evaluation of behaviour there had to be an assessment of behaviour by an observer in a real or simulated environment, using a behavioural observation system. The study had to utilise a pre-post design, and must contain sufficient information to allow Cohen's *d* and the SE to be calculated.
- *Clinical care outcomes*: For an evaluation to be coded as at this level it must include a measure of safety or quality. The assessment must involve a pre-post evaluation and allow Cohen's *d* and the SE to be calculated.

Overall, the screening process resulted in nine separate evaluations of the impact of training on reactions, three evaluations of knowledge change, nine evaluations of attitude change, seven assessments of behaviour change, and seven assessments of changes in clinical care outcomes (see Figure 1).

### **Data extraction process**

Using a data abstraction form with inclusion and exclusion criteria, two authors (POC and AOD) extracted data from included articles. The form collected the following information: year study was conducted, study design, sample size, description of study participants, description of intervention, the nature of the evaluations carried out and the instruments used, the delay between training and evaluation, methodological quality, and sample-size calculation.

### **Quality assessment**

The quality of the papers included in the meta-analysis was evaluated using the 10-item medical education research study quality instrument (MERSQI).<sup>18</sup> The MERSQI was designed to measure the methodological quality of experimental, quasi-experimental, and observational studies in medical education. The instrument assesses quality based on the following dimensions: study design, sampling, type of data, validity of evaluation instrument, data analysis, and outcomes. The MERSQI has been found to have acceptable levels of reliability and validity.<sup>18</sup>

### **Quantitative data synthesis**

The data was analysed separately for each of the four levels of training evaluation or outcome levels. The MIX 2.0<sup>19</sup> software was used to carry out the meta-analyses. A random effects model as utilised by MIX 2.0 was used to calculate the overall effect size and heterogeneity estimates. Effect direction was reversed when necessary to ensure that positive effects indicate an improvement in scores on the measures. Cohen's *d* was used to calculate the effect size. Cochran's *Q*-test was used to assess whether the studies were homogeneous. A significant *Q* is associated with differences between the effects contributing to the mean. In contrast, a nonsignificant *Q* indicates a lack of substantial differences between the effects

contributing to the mean, once sampling error has been removed. Where multiple evaluations were included from the same study (Clay-Williams et al<sup>20</sup> at the attitude level, and Shapiro et al<sup>21</sup> at the behavioural level), they were entered into the meta-analysis independently.

## **RESULTS**

### **Study characteristics**

A total of 20 articles met the a-priori criteria for inclusion in the meta-analysis. The samples include cross-disciplinary teams such as theatre teams, intensive care unit teams, emergency department teams, neonatal unit teams, labour and delivery personnel, operating room personnel (see Table 1). The teams tend to be multi-level consisting of resident physicians, anaesthetists, nurses, midwives, and students. Training time ranges from 90 minutes to 2 days and included didactic lectures, workshops, simulation, coaching. Course content included assertiveness, situation awareness, teamwork, communication, task management, stress management, team coordination, error and crisis management (see Table 1).

Table 1. Description of studies included in meta-analysis.

Author	Year	Training description	Training time	Population
Aebersold et al <sup>22</sup>	2013	CRM training with a focus on assertiveness	Didactic workshop: 6 hrs Simulation: 2 hrs	63 nursing students
Brock et al <sup>23</sup>	2013	Team STEPPS & simulation training	Didactic workshop: 1 hrs Simulation: 3 hrs.	306 medical, nursing, pharmacy, and physician assistant students
Capella et al <sup>24</sup>	2010	Team STEPPS & simulation training	Lectures: 4 hrs Simulation: 2 hrs	91 members of a trauma team
Clay-Williams et al <sup>20</sup>	2013	Adaptation of military CRM for healthcare	Lectures: 1 day	157 doctors, midwives, and nurses
		Adaptation of military CRM for healthcare and simulation	Lectures: 1 day Simulation: 1 day	
France et al <sup>25</sup>	2005	CRM training based upon an aviation model	1 day	182 trauma, emergency & OR staff
Hansel et al <sup>26</sup>	2012	CRM training with a focus on situation awareness	1.5 days	61 final-year medical students
Hicks et al <sup>27</sup>	2012	Nontechnical skills training and simulation	1 day	14 residents
Holzman et al <sup>28</sup>	1995	Anaesthetic CRM training with simulation	6 hrs	72 anaesthetists and nurse anaesthetists
Jankouskas et al <sup>29</sup>	2007	CRM training in teamwork, communication, task management, leadership	3 hrs	40 multi-disciplinary residents and nurses
Kurrek& Fish <sup>30</sup>	1996	Anaesthetic CRM training with simulation	1 day	59 anaesthetists
McCulloch et al <sup>31</sup>	2009	Didactic and interactive teaching based on an aviation CRM model	9 hrs (+ coaching for 3 months)	54 members of theatre staff
Meurling et al <sup>32</sup>	2013	Training in teamwork under stress	Seminars: 4 hrs Simulation: 1 day	151 intensive care unit staff
Morey et al <sup>33</sup>	2002	Emergency team coordination.	8 hrs	1,058 clinical staff
O'Connor et al <sup>34</sup>	2013	Didactic and interactive teaching based on an aviation CRM model	90 mins	110 Irish interns
Reznek et al <sup>35</sup>	2003	Didactic session on human error and crisis management followed by a simulated crisis	Lectures: 90 mins Simulation: 90 mins	13 emergency medicine Residents
Robertson et al <sup>36</sup>	2010	Team STEPPS- lectures followed by small group team exercise	4 hours	213 medical & nursing students
Sawyer et al <sup>37</sup>	2013	Team STEPPS course	Testing & simulation: 2hrs Lectures: 4 hrs	42 neonatal physicians, nurses, & respiratory therapists
Shapiro et al <sup>21</sup>	2004	Didactic and interactive CRM training.	CRM: 8 hrs	4 attending physicians, 4 residents, and 12 emergency department nurses
		Didactic and interactive CRM training, and simulation training	CRM: 8 hrs Simulation: 8 hrs	
Shea-Lewis <sup>38</sup>	2009	CRM type training	Not provided	Labour and delivery personnel (no numbers provided).
Watts et al <sup>39</sup>	2010	Medical team training	1 day	1025 operating room personnel

## Study quality

The quality of the included studies was assessed using the MERSQI. Each article can receive a maximum of three for each of six domains, and a maximum total MERSQI score of 18. Table 2. shows the domain scores and total MERSQI score for all included studies. The studies scored highly on the quality of data analysis (M=2.80, SD= 0.41), quality of outcome assessment (M=2.15, SD= 0.95) and sampling quality (M=2.00, SD= 0.58). Lower scores were for the domains of design (attributable to the inclusion of pre- post designs; M= 1.53, SD= 0.62), type of data (M= 1.70, SD= 1.49), and outcomes (M= 2.15, SD= 0.95). Total mean score for all included studies was 11.63 (SD= 3.00).

Table 2. Study quality assessment MERSQI Domain scores.

<b>Domain</b> (max=3 for each domain)	<b>Mean (sd)</b>
Design	1.53 (0.62)
Sampling	2.00 (0.58)
Type of data	1.70 (1.49)
Validity	1.45 (0.83)
Data analysis	2.80 (0.41)
Outcomes	2.15 (0.95)
<b>Total (max=18)</b>	<b>11.63 (3.00)</b>

## Outcome Evaluation

Of the studies included in the meta-analysis, 9 studies evaluated the training at the reactions level, three studies reported an assessed knowledge change, nine reported an evaluation of attitude change, and eight reported an evaluation of behaviour change. Only two studies that assessed clinical care outcomes were included in the meta-analysis, these evaluated the impact of training on length of hospital stay. Other measures of clinical care outcomes are presented but the data was not synthesised here due to the diversity of measures used.

*Reactions.* Nine studies of participants’ reactions to training are included in the meta-analysis. All of these studies involved a post-training paper and pencil survey in the form of a five point Likert Scale (see Table 3). Overall reactions to training were very positive, a weighted combined mean score of 4.25 out of possible maximum mean score of five indicates that participants found the training to be relevant and useful.

Table 3. Mean score for post-training reactions (scores were measured on a 5 point scale from 1 ‘not useful’ to 5 ‘very useful’).

<b>Author</b>	<b>Year</b>	<b>Measure</b>	<b>n</b>	<b>M</b>
Aebersold et al <sup>22</sup>	2013	7 item questionnaire	37	4.53
France et al <sup>25</sup>	2005	11 item questionnaire	63	4.31
Hansel et al <sup>a,26</sup>	2012	13 item questionnaire	17	4.75
Holzman et al <sup>28</sup>	1995	4 item questionnaire	31 (attending)	3.88
Holzman et al <sup>28</sup>	1995	4 item questionnaire	37 (house officers)	4.24
Kurrek & Fish <sup>b,30</sup>	1996	9 item questionnaire	35	4.00
O’Connor et al <sup>34</sup>	2013	5 item questionnaire	100	4.35
Reznek et al <sup>35</sup>	2003	14 item questionnaire	13	4.43
Robertson et al <sup>36</sup>	2010	15 item questionnaire	192	4.18
<b>Combined<sup>c</sup></b>				<b>4.25</b>

<sup>a</sup>This scale was converted from a 6 point scale to a 5 point scale.

<sup>b</sup>The scale of this study was reversed to be consistent with the other studies analysed.

<sup>c</sup>The combined score was calculated by multiplying the number of participants in each study by the mean score for each study, adding the nine values together and dividing by the total number of participants across all of the studies.

*Knowledge.* Three independent studies reported an analysis of knowledge gain as a result of the training (see Table 4). Assessment involves pre- and post-training knowledge assessment using a self-report tool. A meta-analysis including these three studies resulted in a large mean effect size (see Table 5 and the Forest plot in panel A of Figure 2).

Table 4. Measure and summary data for knowledge, attitudes, and behaviour.

Author	Year	Measure	n	d	Delay <sup>a</sup>
<b>Knowledge</b>					
O'Connor et al <sup>34</sup>	2013	11 item multiple choice/short answer knowledge test	90	1.30	3-8 wks
Robertson et al <sup>36</sup>	2010	12 item knowledge test	240	1.04	None
Sawyer et al <sup>37</sup>	2013	TeamSTEPPS learning benchmarks (23 item test)	42	0.84	None
<b>Attitude</b>					
Brock et al <sup>23</sup>	2013	TeamSTEPPS: Teamwork attitudes scale	149	0.26	None or 2 wks
Clay-Williams et al <sup>20</sup> (classroom only)	2013	Safety Attitude Questionnaire (SAQ) score	34	0.05	None
Clay-Williams et al <sup>20</sup> (classroom & sim)	2013	SAQ score	35	0.50	None
McCulloch et al <sup>31</sup>	2009	14 items on teamwork from SAQ	113	0.27	3 mnths
Meurling et al <sup>32</sup>	2013	14 items on teamwork from SAQ	228	0.24	None
O'Connor et al <sup>34</sup>	2013	12 item attitude questionnaire	92	0.30	3-8 wks
Robertson et al <sup>36</sup>	2010	14 item CHIRP scale	218	0.24	None
Sawyer et al <sup>37</sup>	2013	TeamSTEPPS: Teamwork Attitudes Scale	84	0.38	None
Watts et al <sup>39</sup>	2010	14 items on teamwork from SAQ	3600	0.29	8 mnths
<b>Behaviour</b>					
Hansel et al <sup>26</sup>	2012	Situation awareness global assessment tool	38	0.38	None
Hicks et al <sup>27</sup>	2012	Ottawa CRM global rating scale	20	0.17	None
Jankouskas et al <sup>29</sup>	2007	Teamwork behaviours from ANTS	14	1.6	None
McCulloch et al <sup>31</sup>	2009	Teamwork behaviours from NOTECHS	103	0.44	0-6 mnths
Morey et al <sup>33</sup>	2002	Team dimensions rating form	12	0.47	0-4 mnths
Sawyer et al <sup>37</sup>	2013	TeamSTEPPS performance observational tool	42	1.49	None
Shapiro et al <sup>21</sup> (training only)	2004	Team dimensions rating form	16	0.41	< 2wks
Shapiro et al <sup>21</sup> (training & sim)	2004	Team dimensions rating form	16	0.88	< 2wks
<b>Clinical care outcomes</b>					
Capella et al <sup>24</sup>	2010	Length of stay in hospital	439	0.13	0-2 mnths
McCulloch et al <sup>31</sup>	2009	Length of stay in hospital	103	0.32	0-6 mnths
McCulloch et al <sup>31</sup>	2009	Operating time	103	0.06	0-2 mnths
Shea-Lewis <sup>38</sup>	2009	Number deliveries with adverse events and non-adverse events	8,807	0.31	0-2 yrs.
Capella et al <sup>24</sup>	2010	Time to operating theatre	93	0.48	0-2 mnths
Capella et al <sup>24</sup>	2010	Time in emergency department	439	0.01	0-2 mnths
Capella et al <sup>24</sup>	2010	Length of stay in intensive care	161	0.12	0-2 mnths

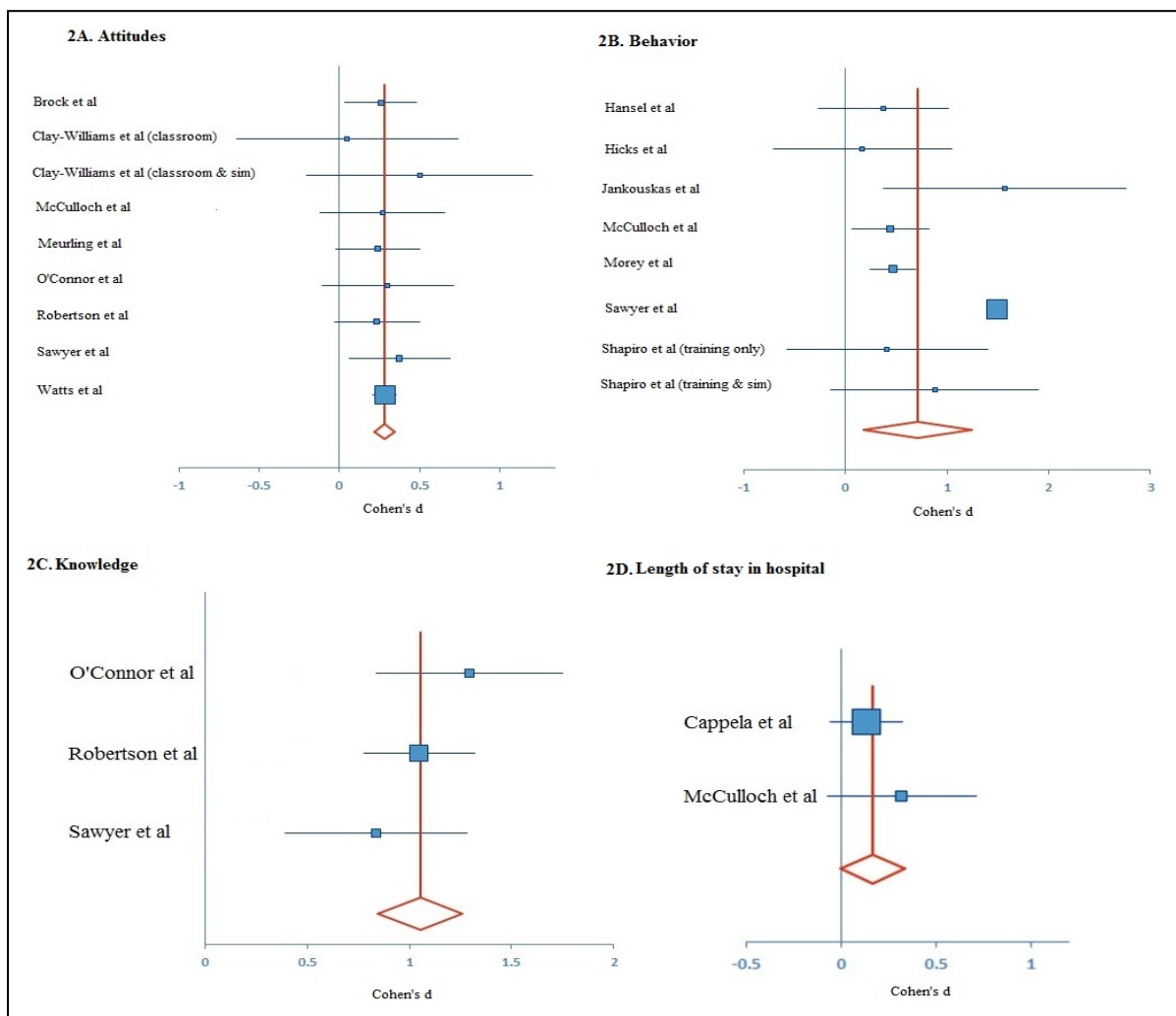
<sup>a</sup>Delay refers to the length of time between completion of the course and the post course evaluation.

Table 5. Meta-analysis results.

Mean effect	K <sup>a</sup>	N <sup>b</sup>	D	95% CI of mean		Q
				Lower	Upper	
Knowledge	3	372	1.05	0.85	1.26	1.99, n.s.
Attitude	9	4,553	0.28	0.22	0.34	1.40, n.s.
Behaviour	8	689	0.72	0.19	1.25	100.36, p<.05
Length of stay	2	552	0.17	0	0.33	0.73, n.s.

<sup>a</sup> the number of studies included in the analysis; <sup>b</sup> the total number of participants in the analysis.

Figure 2. Forest plots for knowledge (panel 2A), attitudes (panel 2B), behaviour (panel 2C), and length of stay in hospital (panel 2D). The zero point on the x-axis is the line of no effect, the size of the square for each study represents the proportion of the weight of the study, the centre line of the diamond is the estimate of the overall effect, and the horizontal width of the diamond is the confidence interval.





*Attitudes.* Nine evaluations of attitude change as a result of training were included in the meta-analysis. Assessments involved questionnaire based pre- and post-assessment of attitudes to teamwork (see Table 4). A meta-analysis of nine different comparisons of attitudes resulted in a small effect size (see Table 5 and the Forest plot in panel B of Figure 2).

*Behaviour.* A total of eight comparisons of the effects on behaviour of the training were included in the meta-analysis. Assessment involved pre- and post-evaluations of team behaviour using behavioural assessment tools (see Table 4). A meta-analysis was carried out, and a large effect size resulted (See Table 5 and the Forest plot in panel C of Figure 2). However, based upon the Q-value; there was significant heterogeneity between the studies (see Table 5).

*Clinical care outcomes.* It was only possible to carry out a meta-analysis of two evaluations of the effect of CRM training on clinical care outcomes; both evaluate the effect of training on length of stay in hospital. No effect was found (see Table 5 and panel D of Figure 2). Other clinical care outcomes measured such as operating time, adverse events, and length of stay in intensive care were also assessed pre- and post-training. However, these evaluations were not included in a meta-analysis due to the diversity of measures, and lack of independence between measures. None-the-less, the effect sizes of these outcome measures are shown in Table 4. The  $d$  statistics for these clinical care outcomes indicate either no effect or a small effect on clinical care outcomes such as time to theatre, time in emergency department or length of stay in intensive care (see Table 4).

## **Synthesis of results**

The meta-analysis of CRM-type team training in healthcare found that participants like this type of training, there was a large effect of the training on knowledge, a small effect of training on attitudes, and a large effect of training on behaviour. The evidence for an effect on clinical care outcomes, or the long term impact of the training was unsupported in this meta-analysis.

## **DISCUSSION**

### **Summary of Evidence**

The primary objective of this meta-analysis was to contribute to the existing literature by quantifying the effects of CRM-type team training in acute care settings at four levels of training evaluation: reactions, learning, behaviour and clinical care outcomes. The evidence presented in this paper shows that participants' reactions to training across studies were overwhelmingly positive. Participants liked the training, and believed that it was relevant to improving teamwork and patient safety. There was a large effect of training on participants' knowledge a small effect on attitudes and a large effect on behaviours. There was insufficient evidence to support an effect on clinical care outcomes or long-term impacts. The findings support the conclusions of previous systematic reviews that report that team training can improve the effectiveness of multidisciplinary teams in acute hospital care.<sup>10-13</sup>

A secondary objective of this analysis was to identify biases in the research evidence in order to improve both the quality of future CRM training interventions and the quality of evaluations of those interventions. Indeed, there are a number of limitations and potential biases in the literature that should be highlighted in order to improve the quality of future research and to aid in the interpretation of the results of this meta-analysis.

*Research design bias.* Overall within the literature, uncontrolled pre-post designs are the most commonly used research design in this domain. Cluster randomised controlled trials and controlled pre and post designs make up about approximately 20% of the literature- depending on the eligibility criteria that is applied. While uncontrolled pre-post designs can produce high quality evidence, they do have a greater risk of unwanted time-related effects and can be influenced by outside social and environmental changes that may affect outcomes of interest.<sup>13</sup> Randomised controlled trials may be the preferable option however standardisation is difficult in such complex environments.

*Sampling bias.* In terms of sampling, little or no information is provided in the articles about how the samples are selected for training. If all participants self-selected, it is possible that these participants are more receptive to such training, than the rest of the population and therefore more likely to benefit from it.

*Publication bias.* It is possible, that training interventions that show no significant effects, are not published, this would serve to present an overestimation of the effectiveness of training. This is a weakness of any systematic review or meta-analysis.

*Selective reporting bias.* The possibility of selective outcome reporting is another limitation of this domain of research. It is not common for research protocols to be published in advance in this area, thus it remains unclear whether selective reporting is an issue in this literature or not.

*Variability in interventions and participants.* Within the literature, there are considerable variations in the length and content of the team training courses. Additionally, studies varied

in terms of the number, specialty, role and seniority of the training participants, these factors can potentially impact the effectiveness of the training intervention.

*Outcome assessment bias.* A lack of standardisation of outcome assessment is a key limitation of the field of research. In effect, a wide variety of sometimes non-validated instruments have been used to measure knowledge, attitude and behaviour change. To illustrate, within the 20 studies included in this meta-analysis, six different measures, or derivatives of measures were used at to assess attitudes and six different scales were used to rate behaviours. In relation to the assessment of behaviours, six of the assessments involved performance on simulated cases,<sup>21, 26, 27, 29, 37</sup> and two assessments involved performance on real medical cases.<sup>31,33</sup> Moreover, two studies reported the assessments of specific teamwork behaviour<sup>29, 31</sup> and the others provided a global assessment of teamwork behaviours.<sup>21, 26, 27, 33,</sup>

37

There is also potential bias in the rigour in which behaviour was assessed in some of the studies. Of the eight evaluations at the behaviour level, examiners were blinded in five of the comparisons<sup>21, 27, 29, 33</sup>, in two evaluations<sup>31, 37</sup> the examiners were not blinded, and in one<sup>26</sup> no information was provided in order to allow a determination to be made. Overall, variability in methods and instruments contributes to the significant heterogeneity found in this meta-analysis.

Moreover, instruments are sometimes inappropriately used in the sense that they measure outcomes that are unrelated to the training received, this is particularly true of some of the attitudes scales used. For example, of the 20 studies included in this meta-analysis, three studies used the Safety Attitude Questionnaire (SAQ). The SAQ measures six patient safety

domains: teamwork climate, safety climate, perceptions of management, job satisfaction, working conditions, and stress recognition.<sup>31, 39</sup> Although the SAQ has good psychometric properties, it is a safety culture survey that addresses attitudes to issues beyond the topics addressed in the training that was provided. To illustrate, the training is unlikely to have an effect on perceptions of management or working conditions. Similarly, Robertson et al<sup>36</sup> used the Collaborative Health Interdisciplinary Relationship Planning (CHIRP) scale. The CHIRP scale has seven sub-scales that address interdependence, recognition, empathy, sharing, dominance, organisational climate and respect. As with the SAQ, this questionnaire examines issues beyond those addressed in the training.

In terms of clinical care outcomes, there was tendency to assess outcomes that were too far removed from the training intervention for any meaningful impact of one on the other to be evident. Outcomes such as length of hospital stay, operating time, length of stay in intensive care are influenced by so many factors, unrelated to training, that it seems unreasonable to anticipate an effect of training on these outcomes. Rather, there is a need for the identification of clinical care outcomes that are more closely related to the training provided. For example, process measures (e.g. introduction of new procedures, use of checklists, assessments of the quality of team briefing) are a useful way to assess the impact of training on unit level processes that may be an intermediary in the path of influence to clinical care outcomes. The situation is summed up succinctly by Zeltser et al<sup>11</sup> who state that “research examining the effectiveness of CRM training will need to operationalise terminology for outcome measures and hone in on specific relationships among variables with explicit plausibility for causation”.(p. 21)

*Inclusion/exclusion bias.* In addition to the limitations associated with the literature in general, the meta-analysis presented here is limited by the small number of studies included in the analyses. The criteria for a meta-analysis are more restrictive than for a systematic review and in this case, have resulted in some well cited studies such as Neily et al<sup>40, 41</sup> and Dunn et al<sup>42</sup> being omitted from the analysis. The exclusion of well cited research papers is common in meta-analysis research. It is well known that the use of meta-analysis techniques leads to a gain in precision at the cost of selectivity.<sup>43</sup> Moreover, every meta-analysis has some inherent bias that can be attributable to the inclusion/exclusion criteria and methods used to research the literature.<sup>44</sup> Nevertheless, the evidence presented in this review concurs with the findings from previous systematic reviews<sup>10-14</sup> regarding the effectiveness of CRM training on reactions, learning and behaviours.

In terms of patient care outcomes, the findings of this meta-analysis may present a more conservative estimate of effectiveness than is presented in previous review studies. For example, Neily et al<sup>41</sup> in their retrospective study found significant reductions in mortality ratios in facilities that had undergone CRM training. Such studies present a stronger case for improvements in clinical care outcomes as a result of CRM training, than were found in this meta-analysis. Similarly, some of the review studies also conclude that team training intervention can improve teamwork processes and patient safety outcomes including mortality and morbidity.<sup>14</sup>

## **Recommendations**

Although the findings from the meta-analysis were positive, these should be considered within the limitations described above. Overall we estimate a moderate to high risk of bias in the included research. Bias can be significantly reduced through the use of standardised

measurement instruments in outcome assessment, ensuring that the instrument is appropriate to the training provided, blinded outcome assessment, representativeness of the sample selected, and using randomised selection where possible. These limitations serve to highlight the need for greater standardisation, more precision and larger scale research projects that are matched carefully to outcome assessment in order to advance the field of research.

Two training programs (TeamSTEPPS and the Medical Team Training- MTT- programme as implemented by the Veterans' Association) appear to offer the best opportunities for standardisation and validation of methods and measure. TeamSTEPPS and MTT are the most widely applied CRM-type team training strategies in healthcare and would appear to offer the best opportunity to conduct the type of large-scale study that is required to assess the impact of CRM training over time. These programmes also include standardised tools for assessing attitudes and behaviours. Therefore, the infrastructure, and tools, for a large-scale impact study already exists.

Although there are limitations in the quality of the research, we believe that there is sufficient evidence to demonstrate the value of CRM type team training in health care. In 2003 Smith and Pell<sup>45</sup> published an article entitled: “parachute use to prevent death and major trauma due to gravitational challenge”. The article illustrates that irrefutable scientific evidence is not necessarily required before new safety practices are adopted, and waiting for this evidence may be unethical.

By way of illustration, the regulators of civil aviation in the UK and US made CRM training mandatory for pilots in 1992<sup>46</sup> and 1998<sup>47</sup> respectively. This decision was based upon a body of evidence supporting CRM training effectiveness that is far weaker than that which

currently exists in the healthcare industry. Indeed, there remains no evidence of the long term impacts of CRM training on aircrew behaviour or system safety, many of the measures used to evaluate aircrew performance and attitudes still require validation,<sup>48</sup> and there is little evidence to support the impact of CRM training on mishap rates- even in the military where mishaps are more frequent than in commercial aviation<sup>49</sup>. Yet there was no wait in either military or commercial aviation for conclusive evidence on the effects of CRM training on mishap rates before mandating training to improve teamwork. The evidence for the effectiveness of CRM training in healthcare is now more robust than that which exists in the aviation CRM literature.

It is undeniable that poor teamwork has a negative impact on patient safety and quality of care, and the evidence consistently shows that CRM training is effective in improving teamwork in healthcare. Yet despite this evidence, the impetus for widespread implementation of CRM training in health care is still largely lacking. “While we may not be able to document statistically the number of lives saved by exemplary CRM practices and lives lost through CRM failures, the evidence for the value of CRM cannot be challenged”(p. 499)<sup>50</sup>. In an era of increased specialisation, more acute and complicated procedures, and shorter hospital stays, the need for effective communication and teamwork is more critical than ever. However, “health care professionals still adhere to few standards for debriefing and communication, and rely on their innate interpersonal skill sets”(p. 14).<sup>11</sup>

## **CONCLUSIONS**

The research reported in this meta-analysis adds to the evidence provided by the growing number of individual studies and systematic literature reviews supporting the effectiveness of



team training in healthcare- at least in the short term- on reactions, knowledge, attitudes, and behaviour. This meta-analysis provides researchers with estimates of the effects of training, and identifies issues that should be avoided when designing CRM evaluation studies.

The long-term effects of CRM training or the impact on clinical care outcomes is not yet established. Gathering this evidence will require more multi-level, multi-centre, multi-specialty, and longitudinal studies. The healthcare industry could wait for this evidence before further implementation of CRM training, or perhaps the ‘parachute approach’ may be more prudent.

#### **FUNDING SOURCE**

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

#### **COMPETING INTERESTS STATEMENT**

None of the authors have any competing interests.

#### **CONTRIBUTORSHIP STATEMENT**

AOD and POC were equal contributors to the manuscript. They carried out the data collection, data analysis, and wrote the draft of the manuscript and rewrote new drafts. IK contributed to the design of the project, data interpretation, provided input to the content of the manuscript and contributed to the revisions of manuscript. All authors read and approved the final manuscript.

## References

1. Flin R, O'Connor P, Crichton M. *Safety at the sharp end: training non-technical skills*. Aldershot, England: Ashgate Publishing Ltd; 2008.
2. Greenberg CC, Regenbogen SE, Studdert DM, *et al*. Patterns of communication breakdowns resulting in injury to surgical patients. *J Am Coll Surg* 2007; 204: 533-40.
3. Kennedy TJ, Regehr G, Baker GR, *et al*. Preserving professional credibility: grounded theory study of medical trainees' requests for clinical support. *BMJ* 2009; 338: b128.
4. Manser T. Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. *Acta Anaesthesiol Scand* 2009; 53:143-151.
5. Lauber JK. Resource Management in the cockpit. *Air Line Pilot* 1984; 53: 20-3.
6. Salas E, Burke CS, Bowers CA, *et al*. Team training in the skies: Does crew resource management (CRM) training work? *Hum Fac* 2001, 43:641-674.
7. Salas E, Wilson KA, Burke CS, *et al*. Does crew resource management training work? An update, an extension, and some critical needs. *Hum Fac* 2006, 48:392-412.
8. O'Connor P, Flin R, Fletcher G. Methods used to evaluate the effectiveness of CRM training: A literature review. *J Hum Fac Aero Saf* 2002; 2: 217-34.
9. O'Connor P, Campbell J, Newon J, *et al*. Crew resource management training effectiveness. A meta-analysis and some critical needs. *Int J Aviat Psy* 2008; 18: 353-68.
10. McCulloch P, Rathbone J, Catchpole K. Interventions to improve teamwork and communications among healthcare staff. *Br J Surg* 2011; 98: 469-79.
11. Zeltser MV, Nash DB. Approaching the evidence basis for aviation-derived teamwork training in medicine. *Am J Med Qual* 2010; 25: 13-23.

12. Buljac-Samardzic M, Dekker-van Doorn CM, van Wijngaarden JD, *et al.* Interventions to improve team effectiveness: a systematic review. *Health Policy* 2010; 94: 183-95.
13. Rabol LI, Ostergaard D, Mogensen T. Outcomes of classroom-based team training interventions for multiprofessional hospital staff. A systematic review. *Qual Saf Health Care* 2010; 19: e27.
14. Weaver SJ, Dy SM, Rosen MA. Team-training in healthcare: a narrative synthesis of the literature. *BMJ Qual Saf* 2014; 23(5): 359-72.
15. Kirkpatrick DL. Evaluation of training. In: Craig RL, Bittel LR, eds. *Training and development handbook*. New York: McGraw Hill 1976; 87-112.
16. Mullen B, Rosenthal R. *Advanced BASIC meta-analysis*. New York, Sage Publications; 1985.
17. Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; 6: e1000097.
18. Reed DA, Beckman TJ, Wright SM, *et al.* Predictive validity evidence for medical education research study quality instrument scores: quality of submissions to JGIM's Medical Education Special Issue. *J Gen Intern Med* 2008; 23(7): 903-7.
19. Bax L. MIX 2.0. *Professional software for meta-analysis in Excel. Version 2.0.1.4*. BiostatXL; 2011.
20. 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: 314-21.
21. Shapiro MJ, Morey JC, Small SD, *et al.* Simulation based teamwork training for emergency department staff: does it improve clinical team performance when added to an existing didactic teamwork curriculum? *Qual Saf Health Care* 2004; 13: 417-21.
22. Aebersold M, Tschannen D, Sculli G. Improving nursing students' communication skills using crew resource management strategies. *J Nurs Educ* 2013; 52: 125-30

23. Brock D, Abu-Rish E, Chiu CR, *et al.* Interprofessional education in team communication: working together to improve patient safety. *BMJ Qual Saf* 2013; 22: 414-23.
24. Capella J, Smith S, Philp A, *et al.* Teamwork training improves the clinical care of trauma patients. *J Surg Educ* 2010; 67: 439-43.
25. France DJ, Stiles R, Gaffney EA, *et al.* Crew resource management training- clinicians' reactions and attitudes. *AORN J* 2005; 82: 213-24.
26. Hansel M, Winkelmann AM, Hardt F, *et al.* Impact of simulator training and crew resource management training on final-year medical students' performance in sepsis resuscitation: a randomized trial. *Minerva Anesthesiol* 2012; 78: 901-9.
27. Hicks CM, Kiss A, Bandiera GW, *et al.* Crisis Resources for Emergency Workers (CREW II): results of a pilot study and simulation-based crisis resource management course for emergency medicine residents. *Can J Emerg Med* 2012; 14: 354-62.
28. Holzman RS, Cooper JB, Gaba DM, *et al.* Anesthesia Crisis Resource Management: real-life simulation training in operating crises. *J Clin Anesth* 1995; 7: 675-87.
29. Jankouskas T, Bush MC, Murray B, *et al.* Crisis resource management: evaluating outcomes of a multidisciplinary team. *Simul Healthc* 2007; 2: 96-101.
30. Kurrek MM, Fish KJ. Anaesthesia crisis resource management training: an intimidating concept, a rewarding experience. *Can J Anaesth* 1996; 43: 430-4.
31. 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): 109-15.
32. Meurling L, Hedman L, Sandahl C, *et al.* Systematic simulation-based team training in a Swedish intensive care unit: a diverse response among critical care professions. *BMJ Qual Saf* 2013; 22: 485-94.

33. Morey JC, Simon R, Jay GD, *et al.* Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MedTeams project. *Health Serv Res* 2002; 37: 1553-81.
34. O'Connor P, Bryne D, O'Dea A, *et al.* Excuse me: teaching junior doctors to speak up. *J Comm J Qual Pat Saf* 2013; 39: 425-30.
35. Reznick M, Smith-Coggins R, Howard S, *et al.* Emergency medicine crisis resource management (EMCRM): pilot study of a simulation-based crisis management course for emergency medicine. *Acad Emerg Med* 2003; 10: 386-9.
36. Robertson B, Kaplan B, Atallah H, *et al.* The use of simulation and a modified TeamSTEPPS curriculum for medical and nursing student team training. *Simul Healthc* 2010; 5: 332-7.
37. Sawyer T, Laubach VA, Hudak J, *et al.* Improvements in teamwork during neonatal resuscitation after interprofessional TeamSTEPPS training. *Neonatal Netw* 2013;32: 26-33.
38. Shea-Lewis A. Teamwork: crew resource management in a community hospital. *J Healthc Qual* 2009; 31: 14-8.
39. 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: 206-9.
40. Neily J, Mills PD, Lee P, *et al.* Medical team training and coaching in the veterans health administration; assessment and impact on the first 32 facilities in the programme. *Qual Saf Health Care* 2010; 19: 360-4.
41. 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.

42. Dunn EJ, Mills PD, Neily J, *et al.*. Medical team training: applying crew resource management in the Veterans Health Administration. *Jt Comm J Qual Patient Saf* 2007; 33: 317-25.
43. Guzzo RA, Jackson SE, Katzell RA. Meta analysis. *Res Org Behav* 1987; 9: 407-42.
44. Rosenthal R, Di Matteo MR. Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annu Rev Psychol* 2001; 52: 59-82.
45. Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ* 2003; 327(7429): 1459-61.
46. Civil Aviation Authority. *Air Operator's Certificates- Operation of Aircraft. CAP 360 Part 1*. Gatwick, England: Author; 1992.
47. Federal Aviation Authority. *Operating Requirements: Domestic, Flag, Supplemental, Commuter, and On- Demand Operation. 14 CFR Parts 119, 121 and 135*. Washington, D.C.: Author; 1996.
48. Helmreich RL, Foushee HC. Why CRM? Emperical and theoretical bases of human factors training. In: Kanki BG, Helmreich RL, Anca J, eds. *Crew resource management*. London, England: Academic Press; 2010; 3-57.
49. O'Connor P, Jones D, McCauley M, Buttrey S. An evaluation of the effectiveness of the U.S. Navy's crew resource management program. *International Journal of Human Factors and Ergonomics* 2012; 1(1): 21-40.
50. Helmreich RL, Anca J, Kanki BG. The future of CRM. In: Kanki BG, Helmreich RL, Anca J, eds. *Crew resource management*. London, England: Academic Press 2010; 494-500.