



Contents lists available at ScienceDirect

Australian Critical Care

journal homepage: [www.elsevier.com/locate/aucc](http://www.elsevier.com/locate/aucc)

## Review paper

## Effectiveness of nontechnical skills educational interventions in the context of emergencies: A systematic review and meta-analysis

María Sánchez-Marco, MSc<sup>a</sup>, Silvia Escribano, PhD<sup>a,\*</sup>, María Rubio-Aparicio, PhD<sup>b</sup>, Rocío Juliá-Sanchis, PhD<sup>a</sup>, María-José Cabañero-Martínez, PhD<sup>a</sup><sup>a</sup> Department of Nursing, Faculty of Health Sciences, University of Alicante, Spain; <sup>b</sup> Department of Health Psychology, Faculty of Social Sciences, University of Alicante, Spain

## ARTICLE INFORMATION

## Article history:

Received 9 October 2022  
Received in revised form  
10 January 2023  
Accepted 11 January 2023

## Keywords:

Crisis resource management  
Critical care  
Emergency medicine  
High-fidelity simulation  
Meta-analysis  
Nontechnical skills

## ABSTRACT

**Introduction:** In recent years, the importance of training healthcare professionals in nontechnical skills using effective methodologies has been increasingly recognised as a means of preventing clinical errors in the practice of health care. The aim of this study was to evaluate the effectiveness of educational interventions on nontechnical skills in the emergency medical services and/or critical care unit settings.**Methods:** A systematic search was carried out in the PubMed, SCOPUS, CINAHL, and Web of Science databases according to predetermined inclusion and exclusion criteria. After the initial search, 7952 records were selected after duplicates removed. Finally, a selection of 38 studies was included for quantitative analysis. Separate meta-analyses of standardised mean changes were carried out for each outcome measure assuming a random-effects model. Cochran's *Q*-statistic and *I*<sup>2</sup> index were applied to verify study heterogeneity. Weighted analyses of variance and meta-regressions were conducted to test the influence of potential moderators and funnel plots using Duval and Tweedie's trim-and-fill method, and Egger's regression test were used to examine publication bias.**Results:** All the variables analysed had a significant effect size, with the exception of situational awareness ( $d_+ = -0.448$ ; 95% confidence interval [CI] =  $-1.034, 0.139$ ). The highest mean effect size was found for knowledge ( $d_+ = -0.925$ ; 95% CI =  $-1.177, -0.673$ ), followed by the mean effect sizes for global nontechnical skills ( $d_+ = -0.642$ ; 95% CI =  $-0.849, -0.434$ ), team nontechnical skills ( $d_+ = -0.606$ ; 95% CI =  $-0.949, -0.262$ ), and leadership nontechnical skills ( $d_+ = -0.571$ ; 95% CI =  $-0.877, -0.264$ ). Similar mean effect sizes were found for attitude ( $d_+ = -0.406$ ; 95% CI =  $-0.769, -0.044$ ), self-efficacy ( $d_+ = -0.469$ ; 95% CI =  $-0.874, -0.064$ ), and communication nontechnical skills ( $d_+ = -0.458$ ; 95% CI =  $-0.818, -0.099$ ). Large heterogeneity among the standardised mean changes was found in the meta-analyses ( $I^2 > 75\%$  and  $p < .001$ ), except for self-efficacy where  $I^2 = 58.17\%$ , and there was a nonstatistical result for Cochran's *Q*. This great variability is also reflected in the forest plots.**Discussion:** The use of simulation interventions to train emergency and critical care healthcare professionals in nontechnical skills significantly improves levels of knowledge, attitude, self-efficacy, and nontechnical skills performance.© 2023 Australian College of Critical Care Nurses Ltd. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Clinical errors arising from the practice of health care significantly affect patient safety, increasing mortality, morbidity, and the financial

costs of healthcare systems.<sup>1</sup> One of the first reports published by the U.S. Institute of Medicine (1999) estimated that 1 million deaths per year were attributable to preventable clinical errors.<sup>2</sup> More recently, a report published by the Organization for Economic Co-operation and Development<sup>3</sup> noted that more than one in 10 patients are harmed by clinical errors during their care, causing more than 3 million deaths per year at an economic cost of between \$1 and 2 trillion annually.<sup>3</sup> A study conducted by the Joint Commission<sup>4</sup> from 2004 to 2015 identified that the three leading causes of these clinical errors were mainly human factors in general, followed by lack of communication and

\* Corresponding author at: Carretera de la Universidad, s/n, 03690, San Vicente del Raspeig, Alicante, Spain. Tel.: +34 96 590 3400 (Ext. 2102).

E-mail addresses: [maria.marco7@ua.es](mailto:maria.marco7@ua.es) (M. Sánchez-Marco), [silvia.escribano@ua.es](mailto:silvia.escribano@ua.es) (S. Escribano), [mrubioaparicio@um.es](mailto:mrubioaparicio@um.es) (M. Rubio-Aparicio), [rjulia@ua.es](mailto:rjulia@ua.es) (R. Juliá-Sanchis), [mariajose.cabanero@ua.es](mailto:mariajose.cabanero@ua.es) (M.-J. Cabañero-Martínez).

leadership. Both are embraced in the broad universe of human factors and linked to the professionals' ability to perform nontechnical skills. These nontechnical skills are defined as “cognitive, social, and personal resource skills that complement technical skills and contribute to safe and efficient task performance”<sup>5</sup> and can be classified into seven categories:<sup>6</sup> skills related to situational awareness, decision-making, communication, teamwork, leadership, stress management, and coping with fatigue. Therefore, preventing clinical errors and delivering high-quality care outcomes requires not only high levels of technical knowledge, clinical competence, and good adaptability but also the acquisition and development of nontechnical skills.<sup>7,8</sup> Training healthcare teams in these skills would therefore appear to be essential.

Various training methodologies have been developed for teaching and learning nontechnical skills.<sup>9</sup> These include the so-called Crisis Resource Management (CRM), developed in the aviation field in the late 1970s<sup>10</sup> and extrapolated in the 1990s to services handling crisis situations, such as emergency medical services and critical care units,<sup>11</sup> among others.

The CRM methodology effectively describes and leverages communication and other components of teamwork to improve performance and the safety of care and, ultimately, the quality of care.<sup>12</sup> Through clinical simulation and the subsequent reflective process,<sup>13,14</sup> the CRM methodology offers a stepwise way for participants to overcome individual, team, and contextual pressures in a controlled environment.<sup>15</sup>

The CRM training methodology has been widely used internationally and in the healthcare field for decades.<sup>16</sup> To the best of our knowledge, two meta-analyses have specifically evaluated the effectiveness of CRM training.<sup>17,18</sup> The analysis by O'Connor et al.<sup>18</sup> included 74 studies from different contexts where CRM had been applied, such as the commercial and military aviation industry, the nuclear industry, offshore oil production, commercial shipping, and health care. The authors concluded that the results were insufficient to demonstrate its effectiveness due to a lack of data and recommended a more rigorous evaluation of the methodology and greater access to data and resources. Meanwhile, O'Dea et al.<sup>17</sup> focused on the evaluation of CRM interventions and the training of critical care professionals in particular, finding them to be effective in the short term. Recent systematic reviews have been published which collect and report in depth many important aspects of CRM methodology<sup>19,20</sup> (or interventions with other methodologies that address nontechnical skills<sup>21,22</sup>). Such reviews discuss in depth the design and structure of the interventions, content, coaches, evaluation methods, and the effectiveness of the interventions, among other issues. However, since 2014, no further evidence has been found in the secondary literature on the effectiveness of educational interventions using the CRM methodology (or covering concepts related to those addressed in the CRM of nontechnical skills) at the meta-analytic level. Although the CRM methodology is widely used, there are also interventions in the literature from other theoretical frameworks, such as TeamSTEPS (Team Strategies and Tools to Enhance Performance and Patient Safety) or SBAR (Situation-Background-Assessment-Recommendation), as well as high-fidelity simulation interventions designed exclusively and without following a specific theoretical framework with the aim to train nontechnical skills.<sup>23,24</sup>

The present meta-analysis was therefore conducted to evaluate the effectiveness of educational interventions involving nontechnical skills (communication, leadership, teamwork, situational awareness, decision-making, fatigue, and/or stress management) and targeting healthcare professionals (nursing, medical, and/or emergency health technicians) in critical care units and emergency medical services, both in-hospital and out-of-hospital.

## 2. Methods

This systematic review and meta-analysis was conducted according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA).<sup>25</sup> The work performed was registered on Open Science Framework (OSF) on March 3rd, 2022 (Registration DOI: 10.17605/OSF.IO/5QPGH).

### 2.1. Study selection

Studies were selected according to the following inclusion criteria: (i) An educational intervention aimed at healthcare professionals had been carried out; (ii) the intervention was aimed at medical professionals, nurses, emergency health technicians, and/or paramedics; (iii) the purpose of the intervention was the acquisition of nontechnical skills; (iv) the setting for the intervention was the emergency medical services (in-hospital/out-of-hospital) and/or critical care unit settings; (v) the articles presented the results in terms of a pre–post or postintervention comparison between groups (randomised controlled trial “[RCT]”, controlled clinical trial “[CCT]”, control trial, quasi-experimental).

In addition, the exclusion criteria were as follows: (i) Studies that were not written in English or Spanish, (ii) the intervention was aimed at undergraduate students, and (iii) those that did not provide sufficient statistical data to calculate the effect size.

### 2.2. Search strategies

The search for studies was conducted in January 2021, using a variety of strategies in order to collect as many articles as possible. The search strategies were developed in conjunction with experts in bibliographic searches in the field of health sciences. In the first instance, a search was carried out in the MEDLINE/PubMed, CINAHL, Web of Science, and Scopus electronic databases. As part of the search strategy, the PubMed thesaurus was consulted, using the following Medical Subject Headings (MeSH) terms: “Crew Resource Management”, “simulation training”, “high fidelity simulation training”, “emergency medicine”, “emergency responders”, “emergency nursing”, “emergency medical technicians”, “emergencies”, “critical care”, and “emergency medical services”. The natural language search terms included in the title and/or abstract fields were: “Crew Resource Management”, “Crisis Resource Management”, “CRM”, “non-technical skills”, “simulation training”, “high fidelity simulation training”, “critical care”, “emergency”, “critical care nursing”, and “emergency medical technicians”. The search strategy used for each database is outlined in [supplementary file 1](#).

A manual search was also performed in journals aimed at the training of healthcare professionals with a specific focus on simulation methodology: “*Emergencies*”, “*Simulation in Healthcare*”, “*International Emergency Nursing*”, “*Medical Education*”, “*Medical Teacher*”, “*Advances in Simulation*”, “*BMC Medicine*”, and “*Nurse Education Today*”. No time restrictions were applied in any of these cases, with the aim of being as exhaustive as possible, compiling all existing interventions that meet the established criteria and analysing their effectiveness. Given that simulation methodology has a track record going back several decades and its effectiveness is unlikely to have been impacted by technological advances, no time restrictions were applied to the search. Lastly, the bibliographic references of the articles included were reviewed to find other studies that could be relevant to the meta-analysis but that had not been found using the aforementioned strategies. The process of evaluating the eligibility of the studies was carried out independently by two investigators (M.S.M. and S.E.). Discrepancies were resolved by a third investigator (M.J.C.M.) in order to reach consensus.

### 2.3. Data extraction

A coding protocol was developed to extract the study characteristics, classifying the extracted variables into three categories: substantive, methodological, and extrinsic.<sup>26</sup> The following substantive variables were coded: (i) level of knowledge; (ii) attitude score; (iii) self-efficacy score; (iv) overall nontechnical skills score; (v) score for teamwork-related skills; (vi) leadership skills score; (vii) communication skills score; (viii) score for situational awareness skills; (ix) medical specialty (accident and emergency, helicopter emergency medical service, or critical care unit); (x) type of methodology (passive, interactive, mixed, or unspecified); (xi) theoretical model (CRM, other, unspecified); (xii) type of simulation (high fidelity, low fidelity, none), and (xiii) scenario objectives (CRM [design to train elements of the CRM or specifically mentioned], nontechnical skills, others [design to train skills other than nontechnical skills], unspecified).

The following methodological variables were extracted: (i) study design; (ii) intervention comparison group; (iii) assessment period; (iv) sample size of the pretest experimental group; (v) sample size of the pretest control group; (vi) sample size of the posttest experimental group; (vii) sample size of the posttest control group; (viii) sample size of the follow-up experimental group; (ix) sample size of the follow-up control group; (x) randomisation of the groups; (xi) dropout rate; and (xii) the domains of the quality scale Medical Education Research Study Quality Instrument (MERSQI)<sup>27,28</sup> (see Assessment of Risk Bias of Selected Studies section). Finally, the extrinsic variables coded were as follows: (i) authors; (ii) year; (iii) educational background of the lead author of the study; and (iv) publication status (published or unpublished and whether it has an ISBN or ISSN).

The coding and extraction of these variables was carried out independently by two researchers (M.S.M. and S.E.). Disagreements between the two researchers who coded these characteristics were resolved with a third researcher (M.J.C.M.).

### 2.4. Assessment of risk of bias of selected studies

The quality of the studies was measured using the MERSQI,<sup>27</sup> which was designed to measure the quality of experimental, quasi-experimental, and observational studies, specifically in medical education research. The tool includes 10 items clustered in six domains ("study design"; "sampling"; "data type"; "validity"; "analysis"; "outcome") and a final overall score. These criteria form six domains, each with a maximum score of 3 and a minimum of 0 or 1, that sum to produce a total score that ranges from 5 to 18. It has generally higher interrater reliability (0.68–0.89) and good relationship with the scale Newcastle–Ottawa Scale–Education ( $\rho = 0.60$ ) for studies of simulation-based trained.<sup>28</sup>

The evaluation was performed in pairs, with each researcher independently evaluating the studies, and then discrepancies were resolved by consensus. All participating researchers (M.S.M., S.E., M.R.A., M.J.C.M., and R.J.S.) were previously trained in extracting and evaluating the methodological quality of studies and in the use of the evaluation instrument used.

### 2.5. Computation of effect sizes

For maximum comprehensiveness, this meta-analysis aimed to include studies with and without a comparison group. The unit of analysis was therefore the group, not the study, and the effect size was the standardised mean rate of change. This index was calculated as the pretest–posttest mean and this difference was divided by the pretest standard deviation:  $d = c(m)(\bar{y}_{pre} - \bar{y}_{post})/S_{pre}$ , with  $c(m)$  being a correction factor for small sample sizes.<sup>29</sup> Negative

$d$  values indicated an improvement in the group from the pretest to the posttest. For standardisation, absolute values of  $d$  of around 0.2, 0.5, and 0.8 can be interpreted as small, moderate, and large magnitudes, respectively.<sup>30</sup>

In order to calculate the standardised mean changes when the necessary information (means and standard deviations) was not reported in the study, the authors of the corresponding studies were contacted to request the necessary missing data. In the absence of a reply, effect sizes were calculated using different procedures depending on the information available in the study. For example, conversion equations from significance tests (e.g.,  $t$ -test and Wilcoxon test) and sample size were used.<sup>29</sup> When results were reported by means of sample size, median, range, and/or interquartile range, the means and standard deviations were estimated by different approximation methods.<sup>31</sup>

Finally, effect sizes were calculated separately for each construct evaluated: knowledge, attitude, self-efficacy, global nontechnical skills, teamwork, communication, leadership, and situational awareness.

### 2.6. Statistical analyses

Separate meta-analyses were carried out for the effect sizes of each outcome measure in at least three studies, by assuming a random-effects model. This model involves weighting each standardised mean change by its inverse variance, defined as the sum of the within-study variance and between-study variance, the latter being estimated by restricted maximum likelihood.<sup>32</sup> For each outcome analysed, a forest plot was constructed showing the individual effect sizes, and the mean effect size calculated with a 95% confidence interval (CI) using the method proposed by Hartung.<sup>33</sup> To assess heterogeneity among the individual standardised mean changes, Cochran's heterogeneity  $Q$  statistic and the  $I^2$  index were calculated. If heterogeneity was found and the number of studies for the outcome was at least 10, an analysis of potential moderator variables was performed.<sup>34</sup> For this purpose, mixed-effects models were assumed applying the  $F$  statistic described by Knapp–Hartung for testing the significance of the moderator variable.<sup>35,36</sup> The QW and QE statistics were calculated to assess the model misspecification of the weighted analyses of variance (ANOVAs) and meta-regressions, respectively, together with an estimate of the percentage of variance accounted for by the moderator variable,  $R^2$ .

Finally, publication bias was tested by constructing funnel plots using Duval and Tweedie's trim-and-fill method<sup>37</sup> and Egger's regression test.<sup>38</sup> A statistically significant result for Egger's test ( $p < .10$ ) was evidence of publication bias. We used  $p < .10$  instead of the usual  $p < .05$  because of the lower statistical power of Egger's test with such a small number of studies.<sup>39</sup>

All statistical analyses were carried out with the *metafor* package in R version 3.2.3.<sup>40</sup>

## 3. Results

### 3.1. Characteristics of the included studies

The screening and coding of studies was performed manually. During the screening process, two reviewers independently excluded 7118 studies based on the title and abstract (Fig. 1). After this first screening, 834 full-text studies were reviewed, with 794 studies that did not meet the inclusion criteria being excluded. We contacted the authors of 11 studies that failed to provide sufficient statistical data to calculate the effect size; only one response was obtained ( $k = 10$  studies were discarded for this reason). Finally, two independent coders extracted data from 38 studies (supplementary file 2) in accordance with the study coding manual.

Table 1 contains the main characteristics of the 38 studies included in the review. The total number of participants was 3144 (min. = 7; max. = 684), with an average of 82.74 participants per sample (SD = 124.83). The mean ages of the participants ranged from 21 to 43.43 years (M = 33.59; SD = 6.73), the standard deviations of their ages ranged from 5.70 to 10.75 years (M = 8.69;

SD = 2.13), and the percentage of females ranged from 11.5 to 94.73 (M = 70.79; SD = 21.85).

We analysed 21 articles (55.3%) aimed at emergency health professionals (including disaster or helicopter emergency medical service), 14 articles (39.5%) aimed at critical care professionals, and two (5.3%) articles aimed at both professionals. The predominant

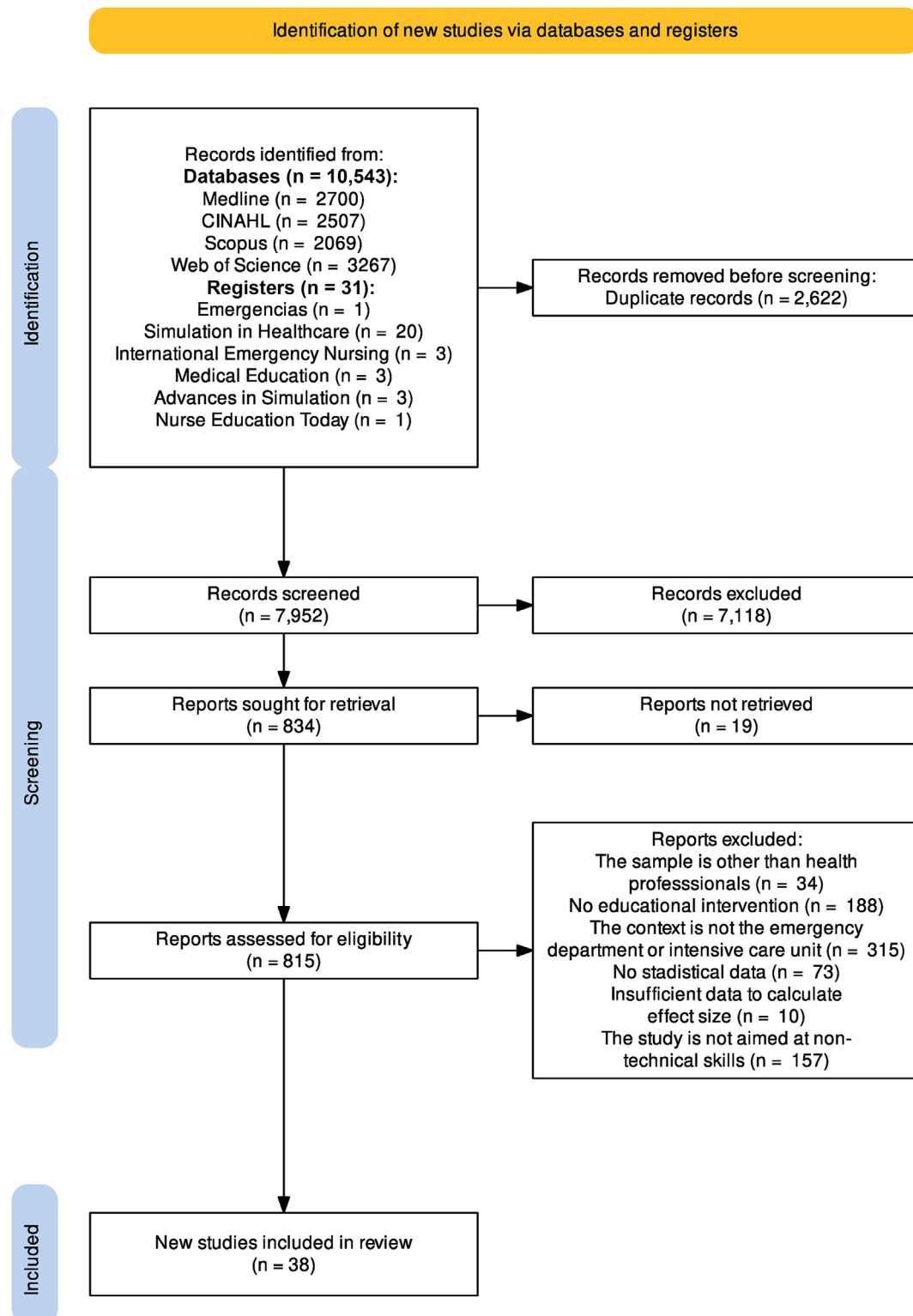


Fig. 1. PRISMA 2020 flow diagram elaborated by Haddaway et al.<sup>53</sup>

**Table 1**  
Descriptive features of 38 studies.

Article	Study design	Outcome measure	Measure tool	Assessment	Year	Country	Sample size	Gender (% female)	Age	Duration, days (hours)	Medical specialty	Health professional	Theoretical model	Type of methodology	Type	Scenarios objectives
Abelsson et al. <sup>1</sup>	Quasi-experimental two groups CG = half of the sessions	1. Global NTS 2. Communication 3. Situational awareness	1. GRS (external observation, validated) 2. GRS communication dimension (validated) 3. GRS situational awareness dimension (validated)	Pretest Posttest (8 weeks)	2017	Sweden	N = 63 EG = 27 CG = 36	Total = 26/ 63 41.27% EG: 12/27 (44.44%) CG: 14/36 (38.89%)	EG: M = 38 (SD = 9, range = 26-63) CC: M = 42 (SD = 9.4, range = 26-63)	4 days	A&E, Catastrophe or HEMS	Nursing	Unspecify	Interactive: Simulation	High-fidelity simulation	Other
Anderson et al. <sup>2</sup>	Quasi-experimental a single group	1. Global NTS	1. Behavioural performance markers (external observation, validated)	Pretest Posttest (immediate)	2006	EEUU	N = 9	8/9 (89%)	M = 32 (range = 26-44)	1 day	Intensive critical care	Nursing	ECMO Sim training program	Mixed: Lectures and simulation	High-fidelity simulation	NTS
Armstrong et al. <sup>3</sup>	Quasi-experimental a single group	1. Global NTS 2. Communication 3. Leadership 4. Situational awareness	1. Modified T_NOTECHS (external observation, validated) 2. T_NOTECHS communication dimension 3. T_NOTECHS leadership dimension 4. T_NOTECHS situational awareness dimension	Pretest Posttest (immediate)	2020	New Zealand	N = 15	-	-	1 day	A&E, Catastrophe or HEMS	Nursing	CRM	Mixed: Lectures and simulation	High-fidelity simulation	CRM
Batchelder et al. <sup>4</sup>	Quasi-experimental a single group	1. Communication 2. Leadership 3. Situational awareness	1. Questionnaire (external observation, no validated) 2. Questionnaire (external observation, no validated) 3. Questionnaire (external observation, no validated)	Pretest Posttest (immediate)	2009	UK	N = 12	-	-	12 days	A&E, Catastrophe or HEMS	Medicine and paramedics	CRM	Mixed: Lectures and simulation	High-fidelity simulation	CRM
Brewster et al. <sup>5</sup>	Quasi-experimental a single group	1. Knowledge 2. Team	1. Questionnaire (self-administered, no validated) 2. Item (self-administered, no validated)	Pretest Posttest (immediate) Follow-up (4 months after)	2017	Australian	N = 99	-	M = 43 (SD = 9.01)	1 day (4 h)	Intensive critical care	Nursing and medicine	CRM	Mixed: Lectures and simulation	High-fidelity simulation	CRM
Burton et al. <sup>6</sup>	Quasi-experimental a single group	1. Knowledge 2. Attitude 3. Global NTS	1. Questionnaire (no validated) 2. Safety Attitude Questionnaire (validated) 3. MPTHS (external observation, validated)	Pretest Posttest (immediate)	2011	EEUU	N = 19	18/19 (94.73%)	-	6 days (24 h)	Intensive critical care	Nursing and medicine	Unspecify	Interactive: Simulation	High-fidelity simulation	NTS
Chamberland et al. <sup>7</sup>	Experiment. CG another intervention without NTS	1. Communication	1. Item (external observation, validated)	Pretest Posttest (immediate) Posttest (3 months after)	2018	Canada	N = 60 EG = 30 CG = 30	24/60 (40%)	M = 40 (SD = 9.3, range = 26-59)	1 day (4 h)	Intensive critical care	Nursing and medicine	CRM	Interactive: Participation	High-fidelity simulation	CRM
Coulours et al. <sup>8</sup>	Quasi-experimental a single group	1. Team 2. Communication	1. SBAR Teamwork component (external observed,	Pretest Posttest (immediate)	2017	EEUU	N = 23 Groups = 5	-	-	4 day (4 h)	A&E, Catastrophe or HEMS	Medicine (residents)	SBAR	Mixed: Lectures and simulation	High-fidelity simulation	NTS

(continued on next page)

Table 1 (continued)

Article	Study design	Outcome measure	Measure tool	Assessment	Year	Country	Sample size	Gender (% female)	Age	Duration, days (hours)	Medical specialty	Health professional	Theoretical model intervention	Type of methodology	Type	Scenarios objectives
Crimlisk et al. <sup>9</sup>	Quasi-experimental a single group	1. Global NTS	no validated) 2. SBAR component (external observation, no validated) 1. Team dynamics and confidence (self-administered, no validated)	Pretest Posttest (immediate)	2017	EEUU	N = 63	—	—	1 day (4 h)	Mixed: A&E, Catastrophe or HEMS and Intensive critical care	Nursing	EART	Mixed: Lectures and simulation	High-fidelity simulation	NTS
Delaney et al. <sup>10</sup>	Quasi-experimental a single group	1. Knowledge	1. TeamSTEPS (self-administered, no validated)	Pretest Posttest (immediate)	2015	EEUU	N = 82	70/82 80.50%	M = 27.6	1 year	Mixed: A&E, Catastrophe or HEMS and Intensive critical care	Nursing	TeamSTEPS	Mixed: Lectures and simulation	High-fidelity simulation	NTS
George et al. <sup>11</sup>	Quasi-experimental a single group	1. Knowledge 2. Team	1. Examen online (no validated) 2. Teamwork skills scale (self-administered, validated)	Pretest Posttest (immediate) Posttest (1month after)	2018	EEUU	N = 22	—	—	1 day (5 h)	Intensive critical care	Nursing and Medicine	Unspecify	Interactive: Simulation	High-fidelity simulation	NTS
Ghazali et al. <sup>12</sup>	Experimental CG less sessions	1. Global NTS 2. Leadership	1. Clinical Teamwork Management (external observation, validated) 2. Behavioural assessment tool (external observation, validated)	Pretest Posttest (immediate)	2019	France	N = 48 EG = 24 CG = 24	22/48 45.83%	—	9 days (12 horas)	A&E, Catastrophe or HEMS	Nursing, Medicine and ambulance driver	Unspecify	Mixed: Lectures and simulation	High-fidelity simulation	NTS
Ginsburg et al. <sup>13</sup>	Quasi-experimental two groups CG = no intervention and intensive critical care professionals.	1. Team	1. Teamwork climate survey (self-administered, validated)	Pretest Posttest (immediate) Follow-up (4 months after baseline)	2017	Canada	N = 71 EG = 42 CG = 29	EG: 35/42 83%	—	Unspecify	A&E Catastrophe or HEMS	Nursing and Medicine	Unspecify	Mixed: Lectures and simulation	Low-fidelity simulation	NTS
Haerlens et al. <sup>14</sup>	Quasi-experimental a single group	1. Attitude	1. Safety attitudes questionnaire (SAQ) (TeamWork climate domain) (validated)	Pretest Posttest (immediate)	2015	The Netherlands	N = 251	—	—	2 days (18 h)	Intensive critical care	Nursing and medicine	CRM	Mixed: Lectures and participative activities	None	—
Hicks et al. <sup>15</sup>	Quasi-experimental a single group	1. Attitude 2. Global NTS	1. Human Factor Attitude Survey (self-administered, validated) 2. Ottawa GRS (external observation, validated)	Pretest Posttest (immediate)	2012	Canada	N = 14	—	—	1 day	A&E Catastrophe or HEMS	Medicine (residents)	CRM	Mixed: Lectures and simulation	High-fidelity simulation	CRM
Houzé-Cerion et al. <sup>16</sup>	Experimental CG, other intervention.	1. Global NTS 2. Team 3. Leadership	1. TEAM Emergency Assessment Measure (external observation, validated) 2. TEAM dimension 3. Team leadership dimension	Pretest Posttest (immediate)	2020	France	N = 211 EG = 105 CG = 106	Total: 145/211 (68.72%) EG: 68/106 (54%) CG: 77/105 (73%)	EG: range = 28-43 CG: range = 26-45	1 day	A&E, Catastrophe or HEMS	Nursing, medicine and other health professionals	CRM	Interactive: Simulation	High-fidelity simulation	CRM
Kanageorge et al. <sup>17</sup>	Quasi-experimental a single group	1. Knowledge 2. Self-efficacy	1. Questionnaire Multiple-choice (self-administered, no validated) 2. Clinical decision-making self-confidence scale	Pretest Posttest (immediate)	2020	EEUU	N = 24	—	—	4 days (4 h)	Intensive critical care	Nursing	Unspecify	Interactive: Simulation	High-fidelity simulation	NTS

Author	Year	Country	Sample Size (N)	Intervention Duration	Setting	Staff	Intervention	Outcome	Fidelity	Simulation Type
Khojrabani et al. <sup>18</sup>	2018	EELUU	N = 13	2 days (10 h)	A&E, Catastrophe or HEMS	Medicine (residents)	Mixed: Lectures and simulation	—	High-fidelity simulation	NTS
Maenhout et al. <sup>19</sup>	2021	Belgium	N = 71	1 day (2 h)	Intensive critical care	Nursing	Interactive: Simulation	67/71 (94.37%)	High-fidelity simulation	NTS
Mah et al. <sup>20</sup>	2009	EELUU	N = 74	1 day	Intensive critical care	Nursing and medicine	Mixed: Lectures and simulation	—	High-fidelity simulation	CRM
Mahramus et al. <sup>21</sup>	2016	EELUU	N = 73	1 day (2 h)	Intensive critical care	Nursing and medicine	Mixed: Lectures and simulation	70%	High-fidelity simulation	CRM
Marzke et al. <sup>22</sup>	2021	EELUU	N = 34	1 day (1 h)	A&E, Catastrophe or HEMS	Nursing and patient care technicians	Mixed: Lectures and simulation	M = 35 (SD = 10.75, range = 22–66)	High-fidelity simulation (real-life)	NTS
Meurling et al. <sup>23</sup>	2013	Sweden	N = 151	2 days (8 h)	Intensive critical care	Nursing and medicine and nurse assistants	Interactive: High-fidelity simulation	119/151 (79%)	High-fidelity simulation	NTS
Morey et al. <sup>24</sup>	2002	EELUU	N = 1058 EC = 684 CG = 374	—	A&E, Catastrophe or HEMS	Nursing, medicine and technicians	Interactive: Participative activities	—	None	—
Munroe et al. <sup>25</sup>	2016	Australian	N = 38	1 day (4 h)	A&E, Catastrophe or HEMS	Nursing	Mixed: Lectures and simulation	M = 29 (range = 21–49)	High-fidelity simulation	NTS
Obernader et al. <sup>26</sup>	2018	EELUU	N = 57	—	A&E, Catastrophe or HEMS	Nursing and medicine	Mixed: Lectures and simulation	—	Low-fidelity simulation	NTS
Parsons et al. <sup>27</sup>	2018	Singapur	N = 14	4 day (5 h)	A&E, Catastrophe or HEMS	Medicine (residents)	Mixed: Lectures and simulation	—	High-fidelity simulation	CRM

(continued on next page)

Table 1 (continued)

Article	Study design	Outcome measure	Measure tool	Assessment	Year	Country	Sample size	Gender (% female)	Age	Duration, days (hours)	Medical speciality	Health professional	Theoretical model intervention	Type of methodology	Type	Scenarios objectives
			dimension 3. Ottawa GRS leadership dimension 4. Ottawa GRS situational awareness dimension													
Pascual et al. <sup>28</sup>	Quasi-experimental a single group	1. Knowledge 2. Global NTS	1. Questionnaire (no validated) 2. Team Leadership-Interpersonal skills (external observation, validated)	Pretest Posttest (immediate)	2011	EEUU	N = 12	8/12 (66.67%)	M = 30.75 (range = 26–36)	3 days (15 h)	Intensive critical care	Nursing	Unspecify	Mixed: Lectures and simulation High-fidelity simulation	High-fidelity simulation	NTS
Patterson et al. <sup>29</sup>	Quasi-experimental a single group	1. Knowledge 2. Attitude	1. Questionnaire (no validated) 2. Safety attitudes questionnaire, SAQ (validated)	Pretest Posttest (immediate) Follow-up (6 months after)	2013	EEUU	N = 289	222/289 (77%)	–	2 days (12 h)	A&E, Catastrophe or HEMS	Nursing, medicine and other health professionals	CRM	Mixed: Lectures and simulation	High-fidelity simulation	CRM
Pennington et al. <sup>30</sup>	Quasi-experimental a single group	1. Global NTS	1. TEAM (external observation, validated)	Pretest Posttest (2–4 weeks)	2018	International (several countries)	N = 10 groups	–	–	12 days	A&E, Catastrophe or HEMS	Nursing and medicine	Unspecify	Interactive	None	–
Peters et al. <sup>31</sup>	Quasi-experimental a single group	1. Knowledge 2. Global NTS 3. Team 4. Communication 5. Leadership	1. TeamSTEPPS (no validated) 2. TTPOI (external observation, validated) 3. TTPOI team dimension 4. TTPOI communication dimension 5. TTPOI leadership dimension	Pretest Posttest (immediate)	2017	EEUU	N = 82	–	–	1 day (8 h)	A&E, Catastrophe or HEMS	Nursing	TeamSTEPPS	Mixed: Lectures and simulation	Low-fidelity simulation	NTS
Pietsch et al. <sup>32</sup>	Quasi-experimental a single group	1. Global NTS	1. Self-evaluation questionnaire (self-administered, no validated)	Pretest Posttest (immediate)	2016	Germany	N = 40	–	–	1 day (1 h)	A&E, Catastrophe or HEMS	Medicine and paramedics	CRM	Mixed: Lectures and simulation	High-fidelity simulation	CRM
Rice et al. <sup>33</sup>	Quasi-experimental a single group	1. Attitude 2. Global NTS 3. Team 4. Communication 5. Leadership	1. T-TAQ (validated) 2. TTPOI (self-administered, validated) 3. TeamSTEPPS Perceptions Questionnaire (self-administered, validated) 4. TTPOI communication dimension 5. TTPOI leadership dimension	Pretest Posttest (immediate)	2016	EEUU	N = 7	6/7 (85.71%)	M = 21	–	Intensive critical care	Nursing	TeamSTEPPS	Interactive: Simulation	High-fidelity simulation	NTS
Ryan et al. <sup>34</sup>	Quasi-experimental a single group	1. Communication 2. Leadership	1. Item (self-administered, no validated) 2. Questionnaire (self-administered, no validated)	Pretest Posttest (immediate)	2018	EEUU	N = 7	–	–	1 day (4 h)	Intensive critical care	Nursing	CRM	Mixed: Lectures and simulation	High-fidelity simulation	CRM
Sauter et al. <sup>35</sup>	Quasi-experimental a single group	1. Knowledge	1. Questionnaire of CRM principles (self-administered, no validated)	Pretest Posttest (immediate)	2016	EEUU	N = 50	34/50 (68%)	Range = 21–60	1 day (7 h)	A&E, Catastrophe or HEMS	Nursing, medicine and other health professionals	CRM	Mixed: Lectures and simulation	High-fidelity simulation	CRM
Shapiro et al. <sup>36</sup>	Experimental CG other intervention.	1. Team	1. BAMS (external observation, validated)	Pretest Posttest (within 2 weeks of the training)	2004	EEUU	N = 20 EC = 2 groups CG = dos groups	–	Range = 29–43	–	A&E, Catastrophe or HEMS	Nursing, medicine and other health professionals	CRM	Mixed: Lectures and simulation	High-fidelity simulation	CRM



Author	Year	Country	N	30/34 (91%)	Range = 25-35	Intensive critical care	Nursing	CRM	Mixed: Lectures and simulation	High-fidelity simulation	CRM
Turkeltown et al. <sup>37</sup>	2017	EEUU	N = 34	30/34 (91%)	Range = 25-35	Intensive critical care	Nursing	CRM	Mixed: Lectures and simulation	High-fidelity simulation	CRM
Winkelmann et al. <sup>38</sup>	2016	Germany	N = 26	3/26 (11.50%)	M = 38.7 (SD = 5.70)	A&E, Catastrophe or HEMS	Medicine	CRM	Mixed: Lectures and simulation	High-fidelity simulation	CRM

**Note:** The articles included are cited in supplementary material Appendix 2.

A&E, emergencies services; BARS, Behaviorally Anchored Rating Scale; CG, control group; CRM, critical resources management; EART, Emergency airway response team; ECMO Sim, Immersive simulation-based training program designed to teach skills to successfully manage life-threatening crises on extracorporeal membrane oxygenation; EG, experimental group; ENNTS, Emergency Nurses' Non-technical Skills; GRS, Global Rating Scale; HEMS, helicopter emergency medical service; HIRAD, History, Identify Red Flags, Assessment, Intervention, Diagnostics, re-assessment and communication; M, average; MPPHS, Mayo High Performance Teamwork Scale; NTS, nontechnical skills; Ottawa GRS, Ottawa Crisis Resource Management Global Rating Scale; SD, standard deviation; SBAR, Situation, Background, Assessment, and Recommendation techniques to transfer the patient to a higher level of care; T, NOTTECHS, Trauma Non-technical Skills teamwork scale; TeamSTEPPS, Team Strategies and Tools to Enhance Performance and Patient Safety; T-TPQ, Team- STEPPS® teamwork and perceptions questionnaire; TTPOJ, Trauma Team Performance Observation Tool; TTAQ, TeamSTEPPS Teamwork Attitudes Questionnaire.

<sup>a</sup> The article by Morey et al.<sup>24</sup> was not included in the meta-analytic analysis.

teaching methodology was the mixed methodology (passive and interactive) ( $n = 27$ ; 71.1%). The predominant theoretical model was CRM ( $n = 18$ ; 47.4%). The most used type of simulation was high fidelity ( $n = 32$ ; 84.2%), and the objectives of the simulation scenarios were mainly geared towards nontechnical skills training ( $n = 18$ ; 47.4%), followed by CRM ( $n = 16$ ; 42.1%), unspecified ( $n = 3$ ; 7.9%), and other ( $n = 1$ ; 2.6%). Most studies ( $n = 32$ ; 84.21%) performed immediate posttest measurements, except for five articles (13.16%) which evaluated the results with variability between 2 and 8 weeks after the intervention, considered as posttest. One article performed the first evaluation measure at 4 months after implementation and, finally, was not included in the statistical analysis. The year of publication of the studies found ranged from 2002 to 2021. In terms of the locations of the studies, four continents were identified: North America, 24 (63.2%); Asia, 1 (2.6%); Europe, 9 (23.7%); Oceania, 3 (7.9%), and International, 1 (2.6%). All studies were written in English.

### 3.2. Assessment of risk bias of selected studies

With respect to the methodological quality of the studies, the average MERSQI<sup>27</sup> score was 12.88 (SD = 2.21), with scores ranging from 8.5 to 16.5 (median = 14) (Table 2). Study designs included randomised experimental studies ( $n = 4$ ; 10.52%) and two-group quasi-experimental designs with pretest and posttest ( $n = 3$ ; 7.89%) and one-group quasi-experimental designs with pretest and posttest (31; 81.58%). These studies predominantly employed a single institution ( $n = 27$ , 71.1%), using objective data by external observer ratings ( $n = 26$ , 68.4%) and assessing knowledge or skills outcomes ( $n = 27$ , 71.1%). Most of the tools ( $n = 28$ , 73.7%) used to assess the outcomes had psychometric tests of internal consistency. All articles (100%) reported statistical inference and were appropriate for the design and types of data collected.

### 3.3. Mean effect size and heterogeneity

Table 3 shows the results of the eight meta-analyses performed for each outcome measure, and the corresponding forest plots are shown in supplementary files 3 and 4. The mean effect sizes calculated for all eight outcomes were negative, indicating an improvement from pretest to posttest. All of them were statistically significant, with the exception of situational awareness ( $d_+ = -0.448$ ; 95% CI =  $-1.034, 0.139$ ), most likely due to the low statistical power of such a small number of studies ( $n = 6$ ).

The highest mean effect size was found for knowledge ( $d_+ = -0.925$ ; 95% CI =  $-1.177, -0.673$ ), followed by the mean effect sizes for global nontechnical skills ( $d_+ = -0.642$ ; 95% CI =  $-0.849, -0.434$ ), team nontechnical skills ( $d_+ = -0.606$ ; 95% CI =  $-0.949, -0.262$ ), and leadership nontechnical skills ( $d_+ = -0.571$ ; 95% CI =  $-0.877, -0.264$ ). Similar mean effect sizes were found for attitude ( $d_+ = -0.406$ ; 95% CI =  $-0.769, -0.044$ ), self-efficacy ( $d_+ = -0.469$ ; 95% CI =  $-0.874, -0.064$ ), and communication nontechnical skills ( $d_+ = -0.458$ ; 95% CI =  $-0.818, -0.099$ ). In all cases, the mean effect sizes were large to medium in magnitude.

Large heterogeneity among the standardised mean changes was found in the meta-analyses ( $I^2 > 75\%$  and  $p < .001$ ), with the exception of self-efficacy where  $I^2 = 58.17\%$ , and there was a nonstatistical result for Cochran's  $Q$  (see Table 3). This great variability is also reflected in the forest plots (see supplementary files 3 and 4).

### 3.4. Analysis of publication bias

With respect to the funnel plots using the trim-and-fill method, for most of the analysed outcomes, the effect sizes were imputed on

**Table 2**  
Quality of studies measured by MERSQI.<sup>27</sup>

Article	Study design	Sampling	Type of data	Validity	Data analysis	Outcome	Total
Abelsson et al. <sup>1</sup>	2	2,5	3	3	3	1,5	15
Anderson et al. <sup>2</sup>	1,5	2	3	2	3	1,5	13
Armstrong et al. <sup>3</sup>	1,5	2	3	3	3	1,5	14
Batchelder et al. <sup>4</sup>	1,5	2	3	1	3	1,5	12
Brewster et al. <sup>5</sup>	1,5	2	1	0	3	1,5	9
Burton et al. <sup>6</sup>	1,5	2	3	3	3	2	14,5
Chamberland et al. <sup>7</sup>	3	2,5	3	2	3	1,5	15
Couloures et al. <sup>8</sup>	1,5	1	3	2	3	1,5	12
Crimlisk et al. <sup>9</sup>	1,5	2	1	1	3	1,5	10
Delaney et al. <sup>10</sup>	1,5	3	1	3	3	1,5	13
George et al. <sup>11</sup>	1,5	2	1	2	3	1,5	11
Ghazali et al. <sup>12</sup>	3	3	3	3	3	1,5	16,5
Ginsburg et al. <sup>13</sup>	2	2	1	3	3	1,5	12,5
Haerkens et al. <sup>14</sup>	1,5	1,5	3	3	3	3	15
Hicks et al. <sup>15</sup>	1,5	2,5	3	3	3	1,5	14,5
Houzé-Cerfon et al. <sup>16</sup>	3	2,5	3	3	3	1,5	16
Karageorge et al. <sup>17</sup>	1,5	2	1	1	3	1,5	10
Khobrani et al. <sup>18</sup>	1,5	2	3	3	3	1,5	14
Maenhout et al. <sup>19</sup>	1,5	2	3	3	3	1,5	14
Mah et al. <sup>20</sup>	1,5	2	3	0	3	1,5	11
Mahramus et al. <sup>21</sup>	1,5	2	3	3	3	1,5	14
Matzke et al. <sup>22</sup>	1,5	2	1	2	3	1	10,5
Meurling et al. <sup>23</sup>	1,5	1,5	3	2	3	3	14
<sup>a</sup> Morey et al. <sup>24</sup>	2	3	3	2	3	3	16
Munroe et al. <sup>25</sup>	1,5	3	3	2	3	1,5	14
Obenrader et al. <sup>26</sup>	1,5	2	1	3	3	1	11,5
Parsons et al. <sup>27</sup>	1,5	2	3	3	3	1,5	14
Pascual et al. <sup>28</sup>	1,5	2,5	3	1	3	1,5	12,5
Patterson et al. <sup>29</sup>	1,5	1,5	3	2	3	1,5	12,5
Pennington et al. <sup>30</sup>	1,5	3	3	3	3	1,5	15
Peters et al. <sup>31</sup>	1,5	2	3	3	3	3	15,5
Pietsch et al. <sup>32</sup>	1,5	2	1	0	3	1	8,5
Rice et al. <sup>33</sup>	1,5	2	3	3	3	1,5	14
Ryan et al. <sup>34</sup>	1,5	2	1	0	3	1	8,5
Sauter et al. <sup>35</sup>	1,5	2	3	0	3	3	12,5
Shapiro et al. <sup>36</sup>	3	2	3	2	3	1,5	14,5
Turkelson et al. <sup>37</sup>	1,5	2	1	2	3	1,5	11
Winkelmann et al. <sup>38</sup>	1,5	2,5	1	0	3	1	9
TOTAL <i>M (SD)</i>	1.87 (0.47)	2.14 (0.45)	2.37 (0.94)	2.03 (1.10)	3 (0)	1.64 (0.57)	12.88 (2.21)

**Note:** All bibliographic references are in [Appendix 2](#).

<sup>a</sup> The article by Morey et al.<sup>24</sup> was not included in the meta-analytic analysis.

the right-hand side of the funnel plots in order to achieve symmetry. For the knowledge measures specifically, three effect sizes were imputed (see [supplementary file 5A](#)), leading to a corrected effect size of  $dc = -0.801$  (95% CI =  $-1.039, -0.564$ ). For the self-efficacy and leadership nontechnical skills measures, two effect sizes were imputed (see [supplementary files 5C and 6C](#), respectively), leading to corrected effect sizes of  $dc = -0.330$  (95% CI =  $-0.548, -0.112$ ) and  $dc = -0.475$  (95% CI =  $-0.757, -0.193$ ), respectively. For global nontechnical skills measures, four effect sizes were imputed (see [supplementary file 5D](#)), leading to a corrected effect size of  $dc = -0.556$  (95% CI =  $-0.748, -0.364$ ). For communication skills, six effect sizes were imputed ([supplementary file 6B](#)), leading to a corrected effect size of  $dc = -0.154$  (95% CI =  $-0.482, 0.174$ ). Finally, for situational awareness measures, one effect size was imputed (see [supplementary file 6D](#)), leading to a corrected effect size of  $dc = -0.326$  (95% CI =  $-0.798, 0.147$ ). However, Egger's test showed statistically nonsignificant results ( $p > .05$ ) for all eight outcome measures, allowing us to rule out publication bias as a threat to the validity of the results of the meta-analyses.

### 3.5. Analysis of moderator variables

The great heterogeneity found among the standardised mean changes led to analysis of the moderator variables. This analysis was applied to those meta-analyses with at least 10 standardised

mean changes, i.e., measures of knowledge, global, team, communication, and leadership skills. [Supplementary files 7 to 12](#) show the results of the weighted ANOVAs and meta-regressions for the results analysed previously.

Of the various potential moderator variables analysed, only simulation showed a statistically significant relationship with the effect sizes on communication nontechnical skills ( $p < .05$ ). To be precise, the mean effect size obtained for high fidelity ( $d_+ = -0.647$ ) was statistically larger than that of low fidelity ( $d_+ = 0.396$ ), with 38.33% of variance accounted for (see [supplementary file 10](#)). Finally, the quality total score did not show a significant effect with any of the outcome measures analysed, i.e., knowledge, global, team, communication, and leadership skills (see [supplementary file 12](#)).

## 4. Discussion

### 4.1. Summary findings

The present meta-analysis was conducted with a view to evaluating the effectiveness of educational interventions that include certain nontechnical skills and are targeted at healthcare professionals in critical care units and emergency departments, both in-hospital and out-of-hospital. Effects were observed on participants' knowledge, teamwork, global nontechnical skills, leadership, attitude, self-efficacy, and communication, assessed in the short

**Table 3**  
Mean effect sizes, 95% confidence, and heterogeneity statistics for the outcome variables.

	n	$d_+$	95% CI		Q	$I^2$
			$d_l$	$d_u$		
Knowledge	13	-0.925	-1.177	-0.673	120.515***	87.67%
Attitude	6	-0.406	-0.769	-0.044	24.880***	81.20%
Self-efficacy	3	-0.469	-0.874	-0.064	4.965	58.17%
Global NTS	19	-0.642	-0.849	-0.434	69.956***	78.00%
Team NTS	13	-0.606	-0.949	-0.262	158.125***	92.48%
Communication NTS	14	-0.458	-0.818	-0.099	104.610***	87.94%
Leadership NTS	10	-0.571	-0.877	-0.264	40.468***	83.15%
Awareness NTS	6	-0.448	-1.034	0.139	24.225***	82.46%

**Note:** NTS, nontechnical skills; n, number of studies;  $d_+$ , Mean effect size; 95% CI, 95% confidence interval for  $d_+$ ;  $d_l$  and  $d_u$ , lower and upper confidence limits; Q, Cochran's heterogeneity Q statistic; with  $k - 1$  degrees of freedom;  $I^2$ , Heterogeneity index; \*\*\* $p < .001$ .

term. There were not enough evaluations to perform a quantitative analysis of the long-term evaluations. O'Dea et al.<sup>17</sup> reflect how most interventions are evaluated through posttest measures and within the first 2 months, concluding that there is no evidence of the long-term impacts of the training.

#### 4.2. Comparison and contrast with previous studies

Unlike previous meta-analyses, our study has focused on specific clinical units where CRM is particularly relevant, such as emergency and critical care units. In the past, the focus was on industrial and/or military contexts, such as the commercial and military aviation industry, air traffic control, the nuclear sector, or commercial shipping.<sup>18</sup> Various medical units were also studied,<sup>17</sup> such as the operating room, trauma, anaesthesiology, obstetrics, and neonatal, as well as emergency and critical care units. Furthermore, while the present meta-analysis sample is of healthcare professionals, including nurses, physicians, and emergency healthcare technicians, others include students and professionals from other disciplines<sup>17,18</sup> with varying degrees of experience, expertise, and training. It is worth noting that this current paper addresses all the dimensions of nontechnical skills,<sup>5</sup> whereas some previous meta-analyses have only focused on specific constructs such as teamwork<sup>41,42</sup> or situational awareness.<sup>43</sup>

With regard to the descriptive characteristics of the studies analysed, in line with previous meta-analyses,<sup>17,18,41,42</sup> most were carried out in the United States. Some of the studies included do not provide theoretical training models to support their interventions, which was already evidenced in the research of Walshe et al.<sup>43</sup> The predominant methodology employed was the mixed methodology,<sup>19</sup> i.e., whereby in addition to being trained by an instructor (passive methodology), healthcare professionals actively participate (interactive methodology) as part of their learning (e.g., training in nontechnical skills using high-fidelity simulation). As the literature indicates, this methodology is more effective than traditional methodologies where training focuses on the transmission of information by an expert in an organised and systematic way, combined with the repetitive practice of clinical procedures and skills.<sup>44</sup> In line with previous meta-analyses,<sup>43</sup> most of the simulations performed were high fidelity. According to Alconero-Camarero et al.,<sup>45</sup> when the sample consists of expert professionals, high-fidelity simulations are more effective. However, if the sample is composed of nonexperts, e.g., nursing or medical students, then medium-fidelity simulations would be more effective because of their level of knowledge and clinical experience.

With regard to the results of the different interventions, significant effect size statistics were obtained for the knowledge, attitude, self-efficacy, global nontechnical skills, teamwork, communication, and leadership variables. In relation to the

outcome variable "knowledge", the effect size observed in our study was larger than that obtained by O'Connor et al.,<sup>18</sup> but smaller than that of the meta-analysis conducted by O'Dea et al.<sup>17</sup> The "attitude" variable also showed an intermediate effect size between that obtained by O'Dea et al.<sup>17</sup> and O'Connor et al.<sup>18</sup> It was not possible to compare the variable "self-efficacy" with previous meta-analyses. However, it is considered a very important variable for the evaluation of training interventions, including nontechnical skills.<sup>46</sup> Self-efficacy refers to healthcare professionals' perception of their abilities in terms of communication skills<sup>47</sup> and is a key predictor of future behaviour.<sup>48</sup> A significant effect size was also observed for "global nontechnical skills" as well as for the specific nontechnical skills measured: teamwork, communication, and leadership. Consistent with these findings, the meta-analysis conducted by McEwan et al.<sup>42</sup> evidenced a moderately large effect size for teamwork. Finally, it should be noted that, as found in previous meta-analyses with healthcare professionals,<sup>43</sup> situational awareness did not show a statistically significant effect with this type of intervention.

However, the high diversity in the tools used for assessing each of the variables should be considered, which could be influencing the heterogeneity found in this meta-analysis, although we have not reported. O'Dea et al.<sup>17</sup> have shown that this diversity is one of the main limitations in this field of research since a wide variety of instruments are used, sometimes not validated, and include different types of measures, such as self-administered and direct behavioural observations. Standardisation of assessment is recommended<sup>17,20</sup> and, on the other hand, to analyse whether the assessment instruments are moderating the results.

In relation to the analysis of moderator variables, only the fidelity level of simulation showed effects on communication nontechnical skills between the team. The results show that interventions are more effective when trained with high-fidelity simulation than those that performed with low fidelity. In keeping with these results, the meta-analysis conducted by Kim et al.,<sup>49</sup> whose aim was to determine the effect size of simulation-based educational interventions in nursing and compare the results according to the fidelity level, identify high-fidelity simulation offers benefits over low-fidelity simulation in cognitive and affective learning outcomes. While the characterised results are of psychomotor type, it evidenced the medium-fidelity simulation (full-body manikins and can be controlled by an external) as more effective. Therefore, adapting the most effective simulation method to meet the proposed results and objectives could be key.<sup>49</sup> In addition, high-fidelity simulation has a high cost. So, it is convenient to assess and consider whether other levels of less expensive simulation can achieve similar results.<sup>50</sup> In any case, it would be advisable to continue research to detect whether there are differences in learning between the different levels.

#### 4.3. Strengths and limitations

This study has some limitations. Firstly, the search strategy may have excluded potential articles, even though various search methods were employed to minimise this possibility. Secondly, most of the studies found were quasi-experimental studies with uncontrolled pretest and posttest designs, thus limiting their internal validity. Randomised controlled trials are the most appropriate option, but they are more complicated to carry out, especially in emergency health contexts. Thirdly, the results regarding the heterogeneity should be interpreted cautiously since the  $I^2$  index presents some limitations. On the one hand, if the number of studies included in the meta-analysis integrates large sample sizes, the  $I^2$  will be larger in the absence of the heterogeneity. On the other hand, with the  $I^2$  being a proportion, the amount of heterogeneity quantified is not absolute.<sup>51,52</sup> Finally, moderate to high levels of heterogeneity were found in the meta-analytic synthesis of all the variables, which influences the final effect of the interventions. Particularly noteworthy is the lack of standardisation in the evaluation of outcomes. A wide variety of instruments, at times not validated, have been used to measure the outcome variables.

Despite these limitations, there are sufficient studies demonstrating that educational interventions, such as CRM, are teaching methodologies that improve nontechnical skills training in the context of emergency departments and critical care units. The use of training programs based on these methodologies to improve the performance of healthcare professionals could improve safety, team performance, and, ultimately, the quality of patient care.<sup>12</sup>

#### 4.4. Implications for educators and clinical practice

Specialised services such as emergency and intensive care require highly trained teams to minimise clinical errors and achieve high-quality care performance, and for this, the acquisition of nontechnical skills such as communication, leadership, and teamwork is critical. The results of this meta-analysis evidence the need for professionals to continue to update their skills throughout their careers, as well as the inclusion of high-fidelity simulation in the curricula of healthcare professionals, especially for the acquisition of the communication skills.

#### 4.5. Areas for future research

A feasible option for further research therefore appears to be the exploration of effective methodologies for situational awareness training, as evidenced in this study. It would also be advisable to conduct more longitudinal, multicentre, and multispecialty studies with the aim of establishing, with adequate methodological quality, the long-term effects of training in these methodologies and the impact of their transfer to clinical practice. It is essential to know if the results obtained are maintained over time<sup>19</sup> and, if not, to know the need to implement reminder interventions and their temporality. All of this will lead to more excellent training for dealing with complex scenarios and, therefore, to improve positive health outcomes, quality of care, and patient safety.

On the other hand, further studies are recommended to examine the efficacy of specific CRM programs versus other simulation programs in other conceptual frameworks aimed at training in nontechnical skills, as well as other factors related to the characteristics of the interventions that could determine their efficacy. The results of this meta-analysis have not been conclusive in this regard for any of the variables studied.

## 5. Conclusions

Learning and teaching nontechnical skills to healthcare professionals in emergency crisis settings using simulation-based methodologies leads to improvements in their knowledge levels, attitudes, and self-efficacy, as well as in the performance of nontechnical skills, both at a global level and in specific skills such as communication, leadership, and work. The situational awareness of the professionals, however, needs to improve.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## CRediT authorship contribution statement

**María Sánchez-Marco:** conceptualisation, methodology, formal analysis, writing -original draft, writing -reviewing & editing and funding acquisition. **Silvia Escribano:** conceptualisation, methodology, formal analysis, writing -original draft, writing -reviewing & editing and funding acquisition. **María José Cabañero-Martínez:** conceptualisation, methodology, resources, reviewing & editing, supervision and funding acquisition. **María Rubio-Aparicio:** reviewing & editing, methodology and formal analysis. **Rocio Juliá-Sanchís:** conceptualisation, methodology, writing -original draft and writing -reviewing & editing and funding acquisition.

## Conflict of interest

The authors state that they have no conflicts of interest.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aucc.2023.01.007>.

## References

- [1] de Montalvo-Jääskeläinen F, Altisent-Trota R, Bellver-Capella V, Cadena-Serrano F, de los Reyes-López M, Gándara-del -Castillo A, et al. Report of the Spanish Bioethics Committee on the ethical aspects of patient safety and, specifically, on the implementation of an effective system for reporting safety incidents and adverse events. 2020. <http://assets.comitedebioetica.es/files/documentacion/Informe%20CBE%20Seguridad%20Paciente.pdf>.
- [2] Kohn LT. To err is human: building a safer health system. National Academies Press; 2000.
- [3] Organization for Economic Cooperation and Development. The economics of patient safety from analysis to action. 2020. <https://www.oecd.org/health/health-systems/Economics-of-Patient-Safety-October-2020.pdf>.
- [4] The Joint Commission. Sentinel event data: root causes by event type 2004-2015. 2016. <https://hcupupdate.files.wordpress.com/2016/02/2016-02-se-root-causes-by-event-type-2004-2015.pdf>.
- [5] Flin Rhona, O'Connor Paul, Crichton Margaret. Safety at the sharp end: a guide to non-technical skills. CRC Press; 2017.
- [6] Leonard M, Graham S, Bonacum D. The human factor: the critical importance of effective teamwork and communication in providing safe care. Qual Saf Health Care 2004;13:i85–90. <https://doi.org/10.1136/qshc.2004.010033>.
- [7] Gordon M, Darbyshire D, Baker P. Non-technical skills training to enhance patient safety: a systematic review. Med Educ 2012;46:1042–54. <https://doi.org/10.1111/j.1365-2923.2012.04343.x>.
- [8] Myers JA, Powell DMC, Psirides A, Hathaway K, Aldington S, Haney MF. Non-technical skills evaluation in the critical care air ambulance environment: introduction of an adapted rating instrument—an observational study. Scand J Trauma Resuscitation Emerg Med 2016;24:24. <https://doi.org/10.1186/s13049-016-0216-5>.
- [9] Jafri F, Mirante D, Ellsworth K, Shulman J, Dadorio N, Williams K, et al. A microdebriefing crisis resource management program for simulated pediatric resuscitation in a community hospital: a feasibility study. Simulat Healthc J Soc Med Simulat 2020;16:163–9. <https://doi.org/10.1097/sih.0000000000000480>.

- [10] Cooper GE, White MD, Lauber JK. Resource management on the flight deck. National Aeronautics and Space Administration; 1980.
- [11] Gaba DM, Fish KJ, Howard SK. Crisis management in anesthesiology. New York [U.A.]: Churchill Livingstone; 1994.
- [12] Pizzi L, Goldfarb NI, Nash DB. Crew resource management and its applications in medicine. In: Dhojanmia KG, Duncan BW, McDonald KM, Wachter RM, editors. Making health care safer: a critical analysis of patient safety practices. Evidence report/technology assessment No 43, AHRQ, vol. 44; 2001. p. 511–9.
- [13] Nader M, Tasdelen-Teker G, DeStephens AJ, Lampotang S, Prelicpean I, Smith RD, et al. Simulation use in outreach setting: a novel approach to building sustainability. *Simulat Healthc J Soc Med Simulat* 2022;17:e136–40. <https://doi.org/10.1097/sih.0000000000000555>.
- [14] Langdalen H, Abrahamsen EB, Sollid SJM, Sørskår LIK, Abrahamsen HB. A comparative study on the frequency of simulation-based training and assessment of non-technical skills in the Norwegian ground ambulance services and helicopter emergency medical services. *BMC Health Serv Res* 2018;18:509. <https://doi.org/10.1186/s12913-018-3325-1>.
- [15] Alinier G. Developing high-fidelity health care simulation scenarios: a guide for educators and professionals. *Simulat Gaming* 2011;42:9–26. <https://doi.org/10.1177/1046878109355683>.
- [16] Gaba DM. Crisis resource management and teamwork training in anaesthesia. *Br J Anaesth* 2010;105:3–6. <https://doi.org/10.1093/bja/aeq124>.
- [17] O'Dea A, O'Connor P, Keogh I. A meta-analysis of the effectiveness of crew resource management training in acute care domains. *Postgrad Med* 2014;90:699–708. <https://doi.org/10.1136/postgradmedj-2014-132800>.
- [18] O'Connor P, Campbell J, Newton J, Melton J, Salas E, Wilson KA. Crew resource management training effectiveness: a meta-analysis and some critical needs. *Int J Aviat Psychol* 2008;18:353–68. <https://doi.org/10.1080/10508410802347044>.
- [19] Buljac-Samardžić M, Dekker-van Doorn CM, Maynard MT. What do we really know about crew resource management in healthcare?: an umbrella review on crew resource management and its effectiveness. *J Patient Saf* 2021;17(8):e929–58. <https://doi.org/10.1097/PTS.0000000000000816>.
- [20] Gross Benedict, Leonie Rusin, Jan Kiesewetter, Zottmann Jan M, Fischer Martin R, Stephan Prückner, et al. Crew resource management training in healthcare: a systematic review of intervention design, training conditions and evaluation. *BMJ Open* 2019;9(2):e025247. <https://doi.org/10.1136/bmjopen-2018-025247>.
- [21] Low XM, Horrigan D, Brewster DJ. The effects of team-training in intensive care medicine: a narrative review. *J Crit Care* 2018;48:283–9. <https://doi.org/10.1016/j.jcrrc.2018.09.015>.
- [22] Buljac-Samardžić M, Doekhie Kirti D, van Wijngaarden JDH. Interventions to improve team effectiveness within health care: a systematic review of the past decade. *Hum Resour Health* 2020;18(1):2. <https://doi.org/10.1186/s12960-019-0411-3>.
- [23] Ounounou E, Aydin A, Brunckhorst O, Khan MS, Dasgupta P, Ahmed K. Nontechnical skills in surgery: a systematic review of current training modalities. *J Surg Res* 2019;76:14–24. <https://doi.org/10.1016/j.jvsurg.2018.05.017>.
- [24] Griffin C, Aydin A, Brunckhorst O, Raison N, Khan MS, Dasgupta P, et al. Non-technical skills: a review of training and evaluation in urology. *World J Urol* 2019;38:1653–61. <https://doi.org/10.1007/s00345-019-02920-6>.
- [25] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. <https://doi.org/10.1136/bmj.n71>.
- [26] Lipsey MW. Identifying interesting variables and analysis opportunities. In: Cooper H, Hedges LV, Valentine JC, editors. *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation; 2009. p. 147–58.
- [27] Reed DA, Beckman TJ, Wright SM, Levine RB, Kern DE, Cook DA. 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 Jul;23(7):903–7. <https://doi.org/10.1007/s11606-008-0664-3>.
- [28] Cook DA, Reed DA. Appraising the quality of medical education research methods: the medical education research study quality instrument and the newcastle-ottawa scale-education. *Acad Med* 2015;90(8):1067–76. <https://doi.org/10.1097/ACM.0000000000000786>.
- [29] Borenstein M, Hedges LV, Higgins JPT. *Introduction to meta-analysis*. West Sussex, England: Wiley; 2009. p. xxix. [https://doi.org/10.1111/j.1751-5823.2009.00095\\_15.x](https://doi.org/10.1111/j.1751-5823.2009.00095_15.x).
- [30] Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ [u.a.]: Erlbaum; 1988.
- [31] Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol* 2014;14:135. <https://doi.org/10.1186/1471-2288-14-135>.
- [32] Cooper HM, Hedges LV, Valentine JC. *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation; 2019.
- [33] Hartung J. An alternative method for meta-analysis. *Biom J* 1999;41:901–16. [https://doi.org/10.1002/\(SICI\)1521-4036\(199912\)41:8%3C901::AID-BIMJ901%3E3.0.CO;2-W](https://doi.org/10.1002/(SICI)1521-4036(199912)41:8%3C901::AID-BIMJ901%3E3.0.CO;2-W).
- [34] Aguinis H, Gottfredson RK, Wright TA. Best-practice recommendations for estimating interaction effects using meta-analysis. *J Organ Behav* 2011;32:1033–43. <https://doi.org/10.1002/job.719>.
- [35] Viechtbauer W, López-López JA, Sánchez-Meca J, Marín-Martínez F. A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychol Methods* 2015;20:360–74. <https://doi.org/10.1037/met0000023>.
- [36] Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med* 2003;22:2693–710. <https://doi.org/10.1002/sim.1482>.
- [37] Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455–63. <https://doi.org/10.1111/j.0006-341x.2000.00455.x>.
- [38] Rothstein HR, Borenstein M, Sutton AJ. *Publication bias in meta-analysis*. New York: Wiley; 2006.
- [39] Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629–34. <https://doi.org/10.1136/bmj.315.7109.629>.
- [40] Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Software* 2010;36:1–48. <https://doi.org/10.18637/jss.v036.i03>.
- [41] Salas E, Diaz Granados D, Klein C, Burke CS, Stagl KC, Goodwin GF, et al. Does team training improve team performance? A meta-analysis. *Hum Factors* 2008;50:903–33. <https://doi.org/10.1518/001872008x375009>.
- [42] McEwan D, Ruissen GR, Eys MA, Zumbo BD, Beauchamp MR. The effectiveness of teamwork training on teamwork behaviors and team performance: a systematic review and meta-analysis of controlled interventions. *PLoS One* 2017;12:e0169604. <https://doi.org/10.1371/journal.pone.0169604>.
- [43] Walshe N, Crowley C, O'Brien S, Browne J, Hegarty J. Educational interventions to enhance situation awareness: a systematic review and meta-analysis. *Simulat Healthc J Soc Med Simulat* 2019;14:398–408. <https://doi.org/10.1097/sih.0000000000000376>.
- [44] Maestre JM, Sztyl D, del Moral I, Ortiz G, Rudolph JW. The training of clinical experts: reflective practice. *Rev Clin Esp* 2013;214:216–20. <https://doi.org/10.1016/j.rce.2013.12.001>.
- [45] Alconero-Camarero AR, Sarabia-Cobo CM, Catalán-Piris MJ, González-Gómez S, González-López JR. Nursing students' satisfaction: a comparison between medium- and high-fidelity simulation training. *Int J Environ Res Publ Health* 2021;18:804. <https://doi.org/10.3390/ijerph18020804>.
- [46] Escribano S, Rocío J, Nereida C, Juana P, Maria José C. Spanish Linguistic validation of the self-efficacy questionnaire in communication skills. *Contemp Nurse* 2021. <https://doi.org/10.1080/10376178.2021.2015415>. ahead-of-print:1-10.
- [47] Chung H, Oczkowski SJW, Hanvey L, Mbuagbaw L, You JJ. Educational interventions to train healthcare professionals in end-of-life communication: a systematic review and meta-analysis. *BMC Med Educ* 2016;16:131. <https://doi.org/10.1186/s12909-016-0653-x>.
- [48] Ajzen I. *From intentions to actions: a theory of planned behavior*. In: *anonymous Action control*. Springer; 1985. p. 11–39.
- [49] Kim J, Park JH, Shin S. Effectiveness of simulation-based nursing education depending on fidelity: a meta-analysis. *BMC Med Educ* 2016 May 23;16:152. <https://doi.org/10.1186/s12909-016-0672-7>.
- [50] Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. *Med Educ* 2012;46:636–47. <https://doi.org/10.1111/j.1365-2923.2012.04243.x>.
- [51] Borenstein M, Higgins Julian PT, Hedges Larry V, Rothstein Hannah R. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Res Synth Methods* 2017;8(1):5–18. <https://doi.org/10.1002/jrsm.1230>.
- [52] Borenstein M. In a Meta-Analysis, the I-squared statistic does not tell us how much the effect size varies. *J Clin Epidemiol* 2022;S0895–4356(22). 00244-X. <https://doi.org/10.1016/j.jclinepi.2022.10.003>.
- [53] Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: an R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Syst Rev* 2022;18:e1230. <https://doi.org/10.1002/cl2.1230>.