Reporting Complex Multidisciplinary Healthcare Team Training Interventions: Development and Validation of a Checklist

Wouter A KEIJSER (✉ wouter@keijser.com)

Universiteit Twente Faculteit Gedrags- Management- en Maatschappijwetenschappen <https://orcid.org/0000-0003-0406-0441>

Robert Wisse

Universiteit Twente Faculteit Gedrags- Management- en Maatschappijwetenschappen

Celeste PM Wilderom

Universiteit Twente Faculteit Gedrags- Management- en Maatschappijwetenschappen

Technical advance

Keywords: Multidisciplinary practice, Teamwork training, Complex intervention, Reporting checklist, TeamSTEPPS

Posted Date: May 7th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-20333/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background Multidisciplinary healthcare teamwork training is regarded as a means to improve patient safety and quality of care. However, it is increasingly characterized by highly complex and contextualized interventions, challenging scholars and practitioners in evaluating reports of such interventions effectively. Conflicting or weak evidence as to whether these complex interventions ultimately affect patient outcomes suggests the need for more comparability, reproducibility, and transparency in scientific reporting. In the absence of an existing reporting guideline, we developed a new checklist for effective reports on the planning, content and implementation design and outcomes of complex teamwork training programs.

Methods The checklist was developed through pragmatically consulting existing relevant frameworks for evaluating complex interventional studies as well as related key literature on teamwork training. We followed a stepped approach in data retrieval, applying a configurative synthesis. Subsequently, the checklist's validity and usability were tested based on a literature review of peer-reviewed journal articles describing the implementation of the widely used comprehensive TeamSTEPPS™ curriculum. We used our checklist to evaluate the reporting in the included articles based on specific selection criteria.

Results The Reporting Complex Multidisciplinary Healthcare Teamwork Training (ReCoMuTe) checklist presented in this paper exists of four distinct categories, each containing components and various explicating elements: Context and preparation; Description; Execution and delivery; Mechanisms of impact. Despite its wide use across various healthcare institutions and settings and the increasing number of publications on its implementation, we found that the reporting in papers on TeamSTEPPS implementation is often inadequate.

Conclusion More standardized scientific reporting on the design and content of complex teamwork training interventions, their contextualized implementation processes and outcomes, can bolster the understanding of effective intervention design, implementation, and evaluation, and can enable well-informed decision making on investments. The ReCoMuTe checklist provides an instrument for evaluating complex interventions, such as studies on TeamSTEPPS implementation, and can assist authors, reviewers, editors, and others in further contemporary research on multidisciplinary healthcare teamwork.

Background

Despite the increasing use of team training (TT) approaches, there is still a dearth of well-defined and comprehensive schemes supported by a wide consensus (Salas et al., 2019). Since around 2000, the number of published academic studies on TT in healthcare has increased dramatically (Fig. 1). Evidence from the last two decades indicates positive associations among elements and effects of TT interventions, even at the level of patient mortality (Buljac-Smaradzic et al., 2010; Weaver et al., 2014; Hughes et al., 2016; Rosen et al., 2018; Salas et al., 2018). However, meta-analytic studies cannot be performed satisfactorily due to the great variability in the 'what,' 'how,' 'when,' and 'who' in the reports on TT, nor can the studied interventions be easily replicated (Cook et al., 2011a; Golub & Fontanarosa, 2015). Notwithstanding the gradual opening up of TT's 'black box,' disclosing various underpinnings of effective and safe teamwork, such as psychological safety and effective team leadership, much of the context of described TT interventions tends to remain unclear. This is because the scholarly publication reports are often inadequate regarding the content, settings, and implementation processes of TT interventions (Salas et al., 2018). Facing the challenge of conveying the particulars of a given TT suitably, we embarked on designing a checklist for such reporting.

Teamwork and Training

Since the publication of 'To Err is Human,' teamwork in healthcare has increasingly gained attention as a serious potential threat to patient safety (Kohn et al., 2000). Over the last 20 years, valuable knowledge and experience has increased in this area, most of which has been derived from aviation and other high-risk industries (Aron & Headrick, 2002; Helmreich et al., 1999; Gross et al., 2019; Rosenbaum, 2019). During this same period, TT has gradually become a standard in quality and safety schemes in healthcare (WHO, 2011; Weaver et al., 2014).

The aim of TT is to mitigate error risks resulting from "natural limitations of human performance and the functioning of complex systems" (Helmreich et al., 1999, p. 29). Trainability of essential healthcare teamwork competencies (i.e., attitudes, behaviors, and cognitions) has been studied extensively (Salas et al., 2005; Suter et al., 2009), and recent meta-analytic reviews provide evidence of the ability of TT to facilitate learning that leads to healthcare professionals' behavioral changes and improved patient and organizational outcomes (Weaver et al., 2014; Hughes et al., 2016; Rosen et al., 2018). Healthcare professionals often participate in various teams, engaging with others under continuously changing circumstances; yet, their individual professional roles mostly remain stable (Andreatta, 2010). Since patient-care pathways occur across many team as well as professional boundaries, effective inter-team and inter-professional communication and collaboration is essential for patient safety. Healthcare settings often involve complex social professional networks with varying backgrounds, educations, viewpoints, work locations, and other characteristics that moderate the effectiveness of their communication and teamwork (Leonard et al., 2004; Marshall et al., 2009). Furthermore, teamwork in healthcare is affected by several characteristics that are distinct from other industries, including low temporal stability and short timespans, a deeply entrenched professional culture (e.g., tacit, complex hierarchical structures), rotating or fluctuating leadership, and protocols that impart high human interdependency (Wildman et al., 2012; Hollenbeck et al., 2012; Reay et al., 2016).

TT in healthcare fosters the improvement of interprofessional collaboration and communication as indispensable elements for (continuous improvement of) the quality and safety of care. TT is characterized by a plurality of teamwork-related topics, skill areas, training settings, instruments, and assessment procedures (Salas et al., 2008; Hughes et al., 2016; Gross et al., 2019). Scholars and practitioners have distinctive streams of knowledge and practices that provide extensive resources, such as crew resource management (CRM), Lean/Six Sigma, and simulation (Koning et al., 2006; Salas et al., 2007; Arora et al., 2015; Gross et al., 2019). Similarly, recent evidence-based training methods and validated measures provide practical instruments for addressing a wide variety of elements that comprise such teamwork (Salas et al., 2013; Marlow et al., 2017). One example is the TeamSTEPPS (i.e., Team Strategies and Tools to Enhance Performance and Patient Safety) curriculum, which was developed as a direct response to the aforementioned 'To Err is Human,' and was developed

by the United States' Agency for Healthcare Research and Quality (AHRQ) and Department of Defense (Kohn et al., 2000; Clancy & Tornberg, 2007; King et al., 2008; Gross et al., 2019).

Numerous moderating factors can facilitate or impede team performance or TT's implementation and sustainability, particularly contextual factors such as organizational conditions and team climates/cultures (Sexton et al., 2000; Clay-Williams & Braithwaite, 2015; West et al., 2015). TT is increasingly couched in the context of other, contemporaneous quality improvement activities geared at large-scale organizational transformation programs (Robichaud et al., 2012; Gross et al., 2019). Hence, TT becomes, through this intertwinement with quality improvement efforts, subject to a wide array of social behavioral, systemic, and context-related factors (Foy et al., 2011). Healthcare TT comprises, in effect, a complex area in which scholars and practitioners alike acknowledge the consequences of re-institutionalization and re-professionalization processes (Reay et al., 2016). Relatedly, recent meta-analytical work revealed the essence of the learning processes of those participating in TT (Hughes et al., 2016). Optimally, healthcare professionals' learning processes follow tactics such as the 'spacing' of series of learning opportunities in near-real life settings (Keijser, 2019). In contrast to 'one-stop-shop' or 'single bullet' training sessions, more longitudinal type-of programs provide openings for such practice-based learning, often combining TT as part of organization quality improvement programs. Moreover, TT sessions that alternate between uni- and multi-professional composition provide learning opportunities in which effective feedback can be shared in psychologically safe settings among peers (Fluger & DeNisi, 1998; Hughes et al., 2016).

Teamwork Training and Complexity

TT should also be placed in the context of complexity thinking that mirrors contemporary healthcare's daily reality in practice and academia, as reflected by the two following considerations. First, healthcare teams function within a context characterized by an ever-increasing fragmentation of members, processes, and technologies, embedded in a myriad of professional systems and structures. This fragmentation may indicate the significance of TT in providing professionals with adequate knowledge, skills, and attitudes for optimal functioning in their fast-paced and fluid network-based interactions and collaborations (Zwarenstein et al., 2009). Second, delivering TT into professionals' relatively chaotic daily (clinical, organizational, and educational) activities also conveys a scholarly challenge: to apply a crosspollination of scientific views to the field of TT (Salas et al., 2018).

Practically, these considerations concur with the tendency of TT-based initiatives to progressively develop into complex, longitudinal programs, often focusing on (continuous) organizational change or practice improvement (Dixon-Woods et al., 2011; Romijn et al., 2019). Scientifically, modern studies on the development of teams (and the multi-team and -unit settings they often occur in), need input from a larger variety of scholarly traditions and viewpoints, also introducing mixed methodologies that can enrich medicine's historical bio-medical approach (Salas et al., 2018). This implies that those reporting TT-based studies should strive for comprehensive (qualitative) descriptions of valuable experiences that are often concealed after (mere quantitative-type) investigations. A coming together of various scholarly streams could provide a better understanding of the efforts, implementations, and wise decision-making regarding TT and the consequent allocation of resources. Contemporary TT-based publications often neglect the importance of implementation contexts by conceptualizing healthcare TT as 'just training.' This disregard might be due to a relative dominance of biomedical over more social science-oriented methodologies in the clinical domain (Greenhalgh, 2012; Pfadenhauer et al., 2017). The present study positions TT as a 'complex TT intervention' (CTTI) (May et al., 2007; Craig et al., 2008).

A complex intervention (CI) refers to implementation or use of various intertwining elements, that are deployed in a complex or 'adaptive' setting or system, and interact by following often unpredictable, non-linear causal processes (Booth et al., 2019; Petticrew et al., 2019). CIs consist "of multiple behavioral, technological, and organizational components" (May et al., 2007, p. 2) and comprises the governing of several interacting components that require distinct expertise and skill from those delivering or receiving the CI. Furthermore, CIs are regularly targeted at several groups, professionals, or organizational levels, and are often tailorable to the context (Craig et al., 2008; Hawe et al., 2009). Settings in which CIs are implemented often represent multi-layered environments comprising various social, cultural, political, procedural, and less-tacit dimensions, jointly structuring the daily reality of the inhabitants (Dixon-Woods et al., 2011). In other words, implementing CIs results in a multifaceted interplay between an intervention's aims, the context in which it is implemented and the various actors that are involved (Petticrew et al., 2019). The causal relationships between a CI and its intended or unexpected outcomes, or the CI's 'active ingredient,' are hard to define, which complicates evaluation. A CI co-exists with its context through an interplay whereby the intervention often requires adaptation to the context, including to changes in the context that occur over time (Øvretveit, 2011).

Various multiprofessional care pathways that encompass mutual decision making and re-organization of care are being designed and studied as CIs (Vanhaecht et al., 2010; Möhler et al., 2015; Moore et al., 2015). In fact, a wide range of quality improvement approaches, including plan-do-study-act cycles, are designed as series of interdependent phases and activities that align with the CI principles; they often comprise changes in social structures (Berwick, 1998; Taylor et al., 2014). Change resistance can play a significant role in such CI implementation; many healthcare professionals can be downright reluctant to participate in improvement initiatives. This holds particularly for professionals with a conservative, biomedical attitude toward adopting new ways of working, such as physicians (Audet et al., 2005; Wilkinson et al., 2011; Keijser et al., 2019). Nevertheless, CIs in which various professionals participate can promote facets of a CTTI (Deneckere et al., 2012). However, researchers should, advisably, explicitly consider whether or not to view an intervention as complex (Booth et al., 2019). Based on the above definitions and the extant use of CIs in healthcare, we contend that most contemporary TT programs can, or should, be epitomized as CTTIs.

Our Aim: Improved Scientific Reporting on CTTIs

The main objective of CI evaluation studies is often to determine whether the goal of the intervention was met. However, how and why an intervention works are equally important questions (Pettigrew et al., 2013; Brewster et al., 2015). Meta-analytical evaluation can contribute significantly to identifying the active ingredients of a CI, as well as the factors impeding its implementation in context. An indicative criterium for appropriate reporting of CTTI evaluation studies could be to judge publications on "whether an expert proficient in [CTTI] research and practice could comprehend and reproduce the intervention based on the information given and the references provided" (Gross, 2019, p. 5). However, Weaver and colleagues reported in their literature review on healthcare TT that "nearly all reviewed studies failed to specify important content" (2010, p. 1756). Failure to describe an intervention's context, design, and outcomes in scientific

publications is also reported in systematic reviews scrutinizing healthcare simulation training, multidisciplinary team and leadership training, and CRM programs (Cook et al., 2011b; Husebo & Akerjordet, 2016; Gross et al., 2019).

Incomplete reporting has been noted in various domains of healthcare, including medical education (Cook et al., 2011a; Glasziou et al., 2014). Yet, “sufficient information must be available for the judgement of the intervention’s clinical benefits, for replication, or for adaption of an intervention to different settings or countries” (Möhler et al., 2013, p. 2). Hence, it is imperative to understand the interplay between intervention and context to facilitate well-informed decisions for choosing, resourcing, delivering, and evaluating a CI adequately (Booth et al., 2019). Compared to conventional controlled trials, evaluating CIs in healthcare is particularly difficult due to the unknown factors and interactions between ‘components.’ Although these components may differ according to the setting, an accurate description of the interventions and their implementation processes as well as the impeding and facilitating factors can contribute to effective reporting (Foy et al., 2011). Currently, much remains unknown about the impact mechanisms and dynamics related to the context and conditions that foster effective application of CTTIs in healthcare (Gross et al., 2019; Salas et al., 2018). Expectedly, multidisciplinary approaches, involving clinicians, scientists, experts, and others at the table, can further this field through a more detailed study of clinical and organizational behavior, the selection, development, and implementation of TT interventions, and their underlying mechanisms of change that accompany the aimed-for outcomes (or lack thereof) (Foy et al., 2011). Theory-driven evaluations can enhance the generalizability and help build a cumulative understanding of the nature of these changes. We believe that framing TT conceptually as a complex intervention accommodates a wider variety of invokable theories and analytic approaches that could govern further scientific work on contemporary TT. After consulting the wealth of extant guidelines and frameworks, we did not find a comprehensive reporting aid, suggesting a need to be addressed. Therefore, we developed a reporting checklist whilst attempting to answer this study’s main research question: *What elements are crucial for adequate reporting of CTTI implementation?*

Methods

Our work was done in two phases: checklist development and face validity testing.

Phase 1: Checklist Development

When developing the checklist, we followed a configurative synthesis of data (Gough et al., 2009; Mohler et al., 2010; Fig. 2). First, we pragmatically inventoried frameworks, guidelines, standards, and approaches relevant to our study’s domain, using online database searches and repositories (e.g., EQUATOR network) (Simera et al., 2010; Bragge et al., 2017). We applied a forward snowballing approach to search for publications using, revising, extending or referring to existing frameworks, and adopted a backward snowballing approach to identify key publications that could provide relevant evidence on the quality of the reporting in the published research (Möhler et al., 2010; Wohlin, 2014). Next, we composed the initial main structure of our envisioned checklist, building on our prior search results. Iteratively, after reviewing retrieved sources and discussions between both researchers, we reached a consensus on the main structure. Then, using the categories framing this initial structure, two authors (WK, RW), individually, identified and recorded relevant sub-items describing elements of reporting and implementing CTTIs. The results were first discussed for relevance and eligibility, and were included after reaching a consensus as elements to the checklist. We continued to search for relevant elements until saturation was reached (i.e., no new aspects were found in the publications). During this development process, we iteratively checked for overlap and redundancy, and also used these feedback loops to ensure clarity in the checklist’s wording. During the entire process, the third author was available for (consensus) discussions; both researchers kept notes and convened at various points to compare findings and discuss their checklist development and to include adequate descriptions of each element.

Phase 2: Face Validity

Testing of the face validity of the 1.0 checklist was based on a literature review of publications reporting the TeamSTEPPS (TS) curriculum. TS is a multifaceted, complex, evidence-based quality improvement intervention, which emphasizes enhancing healthcare team performance through tailorable training designed to enhance team knowledge, skills, and attitudes (Clancy & Tornberg, 2008; Ward et al., 2017). The TS curriculum provides a comprehensive set of connected, but independently utilizable, instruments and approaches that facilitate an understanding and training of key actions for a culture of continuous change based on effective teamwork. TS focuses on such elements as team leadership, mutual performance monitoring, situational awareness, backup behaviors, reflexivity, adaptability, team/collective orientation, shared mental models, mutual trust, and closed-loop communication and feedback (Clancy & Tornberg, 2007; King et al., 2008; Clapper & Kong, 2012; Salas et al., 2018). TS was launched in 2006, and has been implemented in a significant number of various healthcare facilities in and outside the U.S., including: military hospitals (Deering et al., 2011); mental health clinics (Stead et al., 2009); emergency departments (Jones, 2013a); and primary care facilities (Treadwell et al., 2015). Reportedly, TS implementation contributes to improved staff knowledge, skills, and attitudes (Jones et al., 2013b; Sawyer et al., 2013), as well as clinical outcomes (Capella et al., 2010; Spiva et al., 2014).

Our rationale for choosing TS as a feasibility ‘test case’ was twofold. First, it enabled us to review a relatively new intervention, thus deferring a resource-intensive review of a relatively great number of publications featuring a wide variety of approaches. Second, TS is often viewed as an archetypical format for complex multidisciplinary team development interventions (West & Lyubovnikova, 2013; Rosen et al., 2018; Chen et al., 2019; Gross et al., 2019; Hughes et al., 2019), and the variability in its implementation has been suggested to be relevant to understanding its effectiveness (Ward et al., 2017).

Inclusion/exclusion

PubMed, Web of Science, Scopus, Embase, and PsycInfo were consulted to identify reports on TS, using the keywords “TeamSTEPPS” and “Team STEPPS.” Searches were filtered by peer-reviewed journal articles or reviews, published in English, in the period starting from TS’s launch (2007) until November 2017. Eligible publications described TS implementation in multidisciplinary healthcare teams with the goal of improving teamwork and/or patient safety. Studies reporting on TT in interprofessional educational settings, or only describing interventions that were derived from the original TS curriculum (including those

only using a single instrument or tactic, such as huddles or debriefing) were excluded, as were reports of monodisciplinary TS use (e.g., nurses), or quality improvement programs using TS concepts or describing simulation interventions without accounting for the implementation process. To compare implementation methods, we chose studies conducted in Western countries: in Europe, North America, South America, or Oceania. The included literature reviews on TS implementation underwent snowball sampling, to find eligible papers, according to the procedures described by Wohlin (2014). However, no new records were found during this process.

After screening and selecting the abstracts and titles based on our inclusion criteria, full-text versions of the articles were retrieved and reviewed. We excluded reports that a) failed to describe implementation phases or skills that were taught, as well as those that b) did not indicate facilitation by a TS ‘master trainer’ (Clancy & Tornberg, 2008). Since facilitation is imperative to team training (Salas et al., 2002), we regarded the mentioning of a dedicated TS master trainer role as suggestive of well-designed and well-implemented TS programs. A TS master trainer is typified as a person who is knowledgeable and capable of guiding TS implementation, including performing site assessments to determine performance gaps, and coordinating, preparing, training, and providing process consultations at the unit, departmental, and organizational levels (AHRQ, 2014).

Data extraction

Full texts of included articles were reviewed in terms of publication and study characteristics (e.g., country, organizational type, setting, study design, and type of data collection). The selected publications were assessed using Atlas TI™ (version 8.0.33.0; Scientific Software Development GmbH, Berlin, Germany) for text coding based on the identified elements in our reporting checklist. To identify codes and explicatory or exemplary quotations, we initially coded a batch of records thematically (n = 12). Then, we compared and discussed the coded quotes and rated the level of reporting of each publication. Subsequently, both researchers separately coded and rated the remaining article in batches, using the selected quotes from the initial step as a reference. The codes were discussed iteratively, and any disagreements were resolved.

Rating classification of our checklist’s components was done using a 4-point scale: “0” = nothing reported; “1” = minimum reporting: general description, without details; “2” = moderate level of reporting: moderate description or specified report on one or more, but not all of the checklist’s components, and “3” = optimal reporting: specified and detailed reporting. The four scores were also categorized into ‘low’ or ‘high’: “0” or “1” representing a “low” and “2” or “3” representing a “high” score. With a similar method all included publications were also coded for impeding and facilitating factors. Finally, four levels of training evaluation were used to assess each record on how the TT effects were measured (Kirkpatrick, 1994; Jones et al., 2013b).

Results

Checklist Development

In the initial developmental phase, existing evaluation frameworks and reporting checklists, focusing on less linear processes than traditional randomized controlled trials (RCTs), offered valuable input. Our “iterative phased approach that harnesses qualitative and quantitative methods” (Campbell et al., 2000, p. 696) helped to develop our envisioned checklist with its purpose to “assess fidelity and quality of implementation, clarify causal mechanisms and identify contextual factors associated with variation in outcomes” (Craig et al., 2008, p. 34)(Medical Research Council, 2000) and to evaluate “more complex, less advantageous settings” (Glasgow et al., 1999, p. 1322). Eventually, the existing knowledge was imperative for defining our final checklist’s ‘backbone’, comprising four categories: context and preparation; description of intervention; execution and delivery; and mechanisms of impact (Table 1).

Apart from frameworks for the evaluation of worksite-based programs, including those that distinguish facilitating and impeding factors (Fleuren et al., 2004; Wierenga et al., 2016), we sought input from team training interventions targeting healthcare providers (Weaver et al., 2010; Zhu et al., 2015; Marlow et al., 2017). We consulted several frameworks and guidelines describing how to report interventions best, for example the CReDECI 2 guidelines (Criteria for Reporting the Development and Evaluation of Complex Interventions in healthcare), which build on the aforementioned Medical Research Council framework (Campbell et al., 2000; Möhler et al., 2015). During our investigation we also searched for resources focusing on specific study designs, since our main objective was the *reporting* of interventions and their implementation. In addition to the aforementioned sources, we reviewed a broad range of reporting guidelines and frameworks (and their more recent extensions) to find missing components or elements, such as: TIDieR for reporting interventions (Hoffmann et al., 2014); CONSORT and STROBE for simulations (Moher et al., 2010a; Cheng et al., 2016); StaRI for reporting the implementation of CIs (Pinnock et al., 2015; Pinnock et al., 2017); PARIHS for evaluating the implementation of evidence into practice (Rycroft-Malone et al., 2008; Kitson, 2008; Ward et al., 2017); RE-AIM/PRISM for evaluating an implementation (Glasgow et al., 1999; Glasgow et al., 2019); SQUIRE for reporting quality improvement (Goodman et al., 2016); and CReDECI for reporting quality improvement (Möhler et al., 2013; Möhler et al., 2015). These efforts mainly contributed to textual adaptations and alterations, as well as some changes to the composition and ordering of the elements, but did not provide significant additional topics for our checklist.

During the final developmental stages, we formulated components by combining the terminology used in the studied frameworks. Several papers described the prolific growth of terminology in implementation evaluation science and presented guidelines or ‘meta-frameworks’ for them (Damschroder et al., 2009; Bragge, 2017; Pfadenhauer et al., 2017). Therefore, we attempted to use the definitions and terminology provided by these guidelines as often as possible in the Reporting Complex Multidisciplinary Healthcare Teamwork Training (ReCoMuTe) checklist. However, more specific domain-related terminology (i.e., CTTIs) was also included in our final version (Table 1).

Table 1. Categories, components and elements of the ‘Reporting Complex Multidisciplinary Healthcare Teamwork Training’ (ReCoMuTe) checklist

Category	Key components	Elements to describe when reporting
I - Context and Preparation	C1 Needs & barrier assessment	§ Assessment informing (tailored) design, deployment and evaluation § Clear and aligned (e.g. organizational) strategies and justifications § Specification of local problems/challenges
	C2 Engagement & endorsement	§ Motivation, engagement and readiness of the organization and participants (activities and assessments related to, for example: sense-of-urgency; sense-making; shared understanding; coalition formation; resistance reduction; incentives) § Endorsement and support (e.g., management; executives; medical staff) § Clear communication and information (e.g., on program objectives)
	C3 Contextualization	§ Context optimization (e.g., appreciative enquiry envisioned participants; organizational culture change) § Anticipation of simultaneous (possible conflicting) interventions § Intervention adaptations (e.g., design, content, planning) § Feasibility assessment of and pilot testing the conceptual program
	C4 Organization	§ Structure and roles implementation within the organization and members
II - Description of Intervention	D1 Objectives, content, planning & participants	§ Objectives and outcomes* § Participants (e.g., characteristics; selection; recruitment; enrolment); team(s) (e.g., authority differentiation; temporal stability; tasks and skills differentiation) § Content and materials* § Detailed planning activities* (e.g., timing; duration; location; frequency; timeline visualization) § Required resources (e.g., time; finances; materials; location)* § Anticipated causal relation objectives vis-à-vis content and activities* § Metrics on progress and outcomes monitoring* § Theories and evidence (e.g., effectiveness) underpinning the rationale for the design, deployment and evaluation
	D2 Facilitation	§ Planned facilitation strategies, tactics and processes (during and outside planned sessions) § Main facilitator's (e.g., master trainer) characteristics (e.g., background; position; selection; gender) § Additional facilitation (e.g., individual leadership coaching) § Informal facilitation (e.g., support to and from champions) § Resistance/'change fatigue' reduction § Implementation communication (e.g. progress; milestones; updates; materials, such as: pens, badges, (online) nudges)
	D3 Sustainability	§ Sustainment strategies and activity (e.g., integration in organizational quality and educational cycles; periodical evaluation) § Post-intervention (e.g., refresher or new-comers' training)
III - Execution and Delivery	E1 Fidelity / adaptability	§ 'Reach'; dose delivered/ received § Details of delivery activities (temporal; including (un)planned or -intended) § Intervention vis-à-vis context interactions, including (longitudinal) accounts of: (a) (Psychological and other types of) fidelity to intended delivery; (b) adaptations to intervention; (c) changes in context (including those imparted by the intervention) § Contextual detail (e.g., duration; unintended experiences)
	E2 Learning, training & transfer	§ Used training and educational strategies, tactics and methods (e.g., simulation/ didactic sessions; multi-source feedback/ self-reflection; uni-/multidisciplinary approach) § Reception and acceptance of activities (e.g., experiences; responses; satisfaction; feedback)
	E3 Faculty	§ Trainers, coaches, guest lecturers, etc. (e.g., profession; background; (hierarchical) position; motivation; selection; tasks; experiences)
IV - Mechanisms of Impact	M1 Evaluation & analysis	§ Methodology for (process) evaluation (including theoretical rationale) § Descriptions and/or selective narratives reflecting topics relating change management to organization behavioral and other social structures (<i>What happened?</i>) § Impact mechanisms (e.g., anticipated versus unexpected) § Mediating (expected and unexpected) factors, pathways and consequences § Strengths and limitations of intervention's components

- § Lessons learned
- § Possible (causal) explanations of activities' impact on outcomes
- § Description of the control group and conditions (if applicable)

(*at various (organizational) levels)

Face Validity Testing

Our initial search resulted in 1,100 potential abstracts (Figure 2). After removing duplicates (n = 347), the remaining abstracts (n = 753) were screened independently by two researchers (RW and WK), after which they discussed any disagreements. When criteria were not described fully in the title and abstract, or in cases where the researchers doubted the nature of the quality of the improvement, the article was included for full-text analysis. One article (Motley & Dolansky, 2015) in this literature review could not be retrieved. Snowballing reviews did not result in new records. The full text of the papers (n = 135) were then assessed by two researchers (RW and WK) for eligibility, and any disagreements were discussed. A third reviewer (CW) was available for the latter process but was not needed. The entire literature-selection process produced a total of 27 articles for this review (Table 3).

Figure 3. Flowchart of the literature-selection process

Study characteristics

The majority of the included studies originated from U.S. healthcare facilities (n = 24), or U.S.-based organizations such as the U.S. military (n = 1; see, Table 2). Other studies were often conducted in acute or surgical clinical settings (n = 18). Six studies described TS implementation across a hospital system or healthcare region. Additionally, about half of the included studies were conducted in university hospital settings (n = 14), while the remaining studies were performed in mental health, rural hospital, military, or general hospital settings. Two publications did not report the studied TS setting.

Fifteen studies used a pre-post intervention study design. Other study designs used longitudinal (n = 6); mixed methods; cohort study; cluster design experimental study; or cross-sectional comparisons with a pre-post interventional design. One study did not report its research design. Surveys were the most used data collection method (n = 22). Only four studies used qualitative data collection methods, such as interviews, focus groups, or anecdotal evidence. Observational data were used in nine of the 27 studies.

Five studies applied all of Kirkpatrick's four levels of team training assessment as evaluation methods to measure intervention effects (Kirkpatrick, 1994; Table 2). Three articles measured three levels, nine measured two levels, and eight measured one level. Two papers did not report any level of training evaluation.

Quality of reporting

An assessment of the included publications' levels of applying the eleven elements of the ReCoMuTe checklist revealed that a majority of the studies scored low on most of them (Table 3). The most extensively reported were from the checklist's second category, description of intervention, namely the elements 'objectives, content, planning, and participants' (D1) whereby more than half (52%) of the studies had moderate or high scores (Table 3). All the other components were, on average, reported insufficiently, scoring nil or minimum.

Applying the four-level ratings to the E2 ('Learning & Transfer') and E3 ('Faculty') elements was difficult due to the absence or meager reporting details. Therefore, we decided to rate these elements based on a two-tier scoring system: 'reported' or 'not reported' (Table 3). Consequently, the summed ratings for these two elements are less comparable than the others.

The Thomas and Galla (2013) study was optimal on almost all the checklist's components, followed by four others (Stead et al., 2009; Turner, 2012; Brodsky et al., 2013; Jones et al., 2013b).

Table 2. Study characteristics of the included (n=27) studies reporting on I implementation, including their application of Kirkpatrick's levels of training evaluation

Study	Country	Hospital type	Setting	Study design	Data collection method	Kirkpatrick's levels of training evaluation			
						Level 1: reaction	Level 2: learning	Level 3: transfer	Level 4: results
Stead et al. (2009)	Australia	Mental health clinic	Hospital wide	Pre-post training	Surveys/observation	●	●	●	●
Capella et al. (2010)	USA	Academic	Trauma Center	Pre-post training	Observations/clinical outcome data	●	●	●	●
Deering et al. (2011)	USA/Iraq	US Military Healthcare System	Multiple Combat Support Hospitals	Pre-post training	Clinical outcome data	●	●	●	●
Forse et al. (2011)	USA	Academic	OR	Pre-post training	Surveys/clinical outcome data	●	●	●	●
Mayer et al. (2011)	USA	Academic	Pediatric & surgical ICU	Longitudinal	Surveys/interviews/observations/clinical outcome data	●	●	●	●
Mahoney et al. (2012)	USA	Mental health	Hospital wide	Pre-post training	Surveys	●	●	●	●
Turner (2012)	USA	Academic center	ED	Not reported	Anecdotal	●	●	●	●
Brodsky et al. (2013)	USA	Academic center	NICU	Pre-post training	Surveys	●	●	●	●
Jones et al. (2013A)	USA	Multiple hospitals	ED	Pre-post training	Surveys	●	●	●	●
Jones et al. (2013B)	USA	Multiple critical access hospitals	Hospital wide	Cross-sect., comparison; pre-post training	Surveys	●	●	●	●
Sawyer et al. (2013)	USA	Army Medical Center	NICU	Pre-post training	Surveys/observation	●	●	●	●
Sheppard et al. (2013)	USA	Non-profit hospital	Hospital wide	Pre-post training	Observations	●	●	●	●
Thomas & Galla (2013)	USA	Non-profit hospital	Hospital wide	Pre-post training	Surveys	●	●	●	●
Klipfel et al. (2014)	USA	Academic center	Urology surgery	Pre-post training	Surveys	●	●	●	●
Spiva et al. (2014)	USA	Unknown	Acute care	Longitudinal	Surveys/observation	●	●	●	●
Amaya-Añas et al. (2015)	Colombia	Unknown	OR	Pre-post training	Surveys	●	●	●	●
Beitlich (2015)	USA	Rural clinics	L&D and NICU	Pre-post training	Surveys	●	●	●	●
Fischer et al. (2015)	USA	Tertiary military center	OR	Cohort study	Surveys	●	●	●	●
Gupta et a. (2015)	USA	Academic center	Interventional ultrasound	Pre-post training	Surveys	●	●	●	●
Scotten et al. (2015)	USA	Academic center	Pediatric	Longitudinal	Surveys	●	●	●	●
Sonesh et al. (2015)	USA	Academic center	Obstetrics	Mixed methods	Surveys/observations	●	●	●	●
Treadwell et al. (2015)	USA	Primary care	Primary care	Cluster design experimental	Surveys	●	●	●	●
Gaston et al. (2016)	USA	Academic center	Oncology	Mixed methods	Surveys/focus groups	●	●	●	●

Lisbon et al. (2016)	USA	Academic center	ED	Longitudinal	Surveys	●	●	●	●
Rhee et al. (2016)	USA	Academic center	Peri-OR	Longitudinal	Observations	●	●	●	●
Wong et al. (2016)	USA	Academic center	ED	Longitudinal	Surveys	●	●	●	●
Peters et al. (2017)	USA	Academic center	ED	Pre-post training	Surveys/observations/clinical outcome data	●	●	●	●
						9 (33%)	16 (60%)	13 (48%)	17 (62%)

Legend to Table 2

Not reported (●); reported (●)

Percentages represent the number of papers using the Kirkpatrick level for training evaluation compared with the total (n=27).

OR, Operation Room; ICU, Intensive Care Unit; ED, Emergency Department; NICU, Neonatal Intensive Care Unit; L&D, Labor and Delivery.

Table 3. Scoring of the publications on TeamSTEPPS implementation (n=27) with the ReCoMuTe checklist components, including summated scores

Author	1	2	3	4	5	6	7	8	9	10	11	12	13	14					
year																			
	Stead et al. (2009)	Capella et al. (2010)	Deering et al. (2011)	Forse et al. (2011)	Mayer et al. (2011)	Mahoney et al. (2012)	Turner (2012)	Brodsky et al. (2013)	Jones et al. (2013A)	Jones et al. (2013B)	Sawyer et al. (2013)	Sheppard et al. (2013)	Thomas & Galla (2013)	K...e (2013)					
Code																			
C1	●	●	●	●	●	●	●	●●	●	●●	●	●	●●	●					
C2	●●	●	●	●	●	●●●	●●	●●	●	●●	●	●●	●●●	●					
C3	●●	●	●	●	●●	●	●●●	●	●	●	●	●	●●●	●					
C4	●	●	●	●●	●	●●●	●	●●	●	●	●	●●	●●●	●					
D1	●●	●●	●●●	●	●●	●●	●	●●●	●	●●	●●●	●	●●●	●					
D2	●●●	●	●	●	●	●	●	●	●	●	●	●	●●●	●					
D3	●	●	●	●	●	●●●	●●	●●	●	●●●	●	●	●●●	●					
E1	●	●	●●	●	●	●	●●●	●	●	●	●●	●	●●	●					
E2 *	□	■	□	□	□	■	□	■	□	□	□	■	■	■					
E3 *	■	■	■	■	■	■	■	■	■	■	□	■	■	■					
M1	●	●	●	●●●	●	●	●●	●●●	●	●●●	●●	●	●●●	●					
Low score	44%	89%	78%	78%	56%	44%	33%	89%	44%	67%	78%	0%	67%	89%	78%	89%	100%	78%	89%
(0 or 1)																			

High score	56%	11%	22%	22%	22%	44%	56%	67%	11%	56%	33%	22%	100%	33%	11%	22%	11%	0%	22%	11%
(2 or 3)																				

Legend to Table 3

Note the Table Four-level rating on ‘reporting level’: ‘nil’ (●), ‘minimum’ (●●), ‘moderate’ (●●●), ‘high’ (●●●●).

*Components E2 (Learning, training & transfer) and E3 (Faculty) scored with two-level rating: ‘not reported’ (□), ‘reported’ (■).

Low score percentages represent total scores on reporting levels ‘nil’ (●) or ‘minimum’ (●●); high score percentages represent total scores on reporting levels ‘moderate’ (●●●), ‘high’ (●●●●).

Cells with summated percentages scores > 50 are accentuated with **bold** text.

Final version

During the checklist’s development as well as validity evaluation, while assuring no element was left out and the meaning of the descriptions remained intact, we searched for an overlap between the elements and for well-balanced appropriateness, clarity, and conciseness in wording. While validity testing helped to augment the elements’ description conciseness, it did not reveal new categories, components, or elements for our checklist, suggesting completeness in our design for assessing reports on the implementation of TS or similar CTTIs.

I - Context & preparation (C1-4)

The first category of the checklist (see, Table 1) describes the context and preparatory activities before actual deployment of an intervention. The environment or setting in which the intervention is implemented and deployed is referred to as ‘context’ (Pfadenhauer et al., 2017; Booth et al., 2019). Further, the checklist sets out to address four questions: *Did the implementers asses the needs and requirements?* (C1) *Is there support from leadership as well as from participants?* (C2) *How is the training intervention contextualized?* (C3) *How and by whom is the intervention organized?* (C4)

The local infrastructure, availability of resources, knowledge, and experience can vary significantly between healthcare organizations, subsequently affecting implementation potential and execution (Kitson et al., 2008). Moreover, the significance of barrier assessments to inform anticipatory sustainability efforts and the involvement of decision-makers and management have been discussed in previous studies (Kemper et al., 2014; Moffatt-Bruce et al., 2017). Adapting interventions to contextual settings has been described as a tactic to create, for example, shared ownership of improvement among staff (Haerkens et al., 2018).

II – Description of intervention (D1-3)

The second category provides an overview of the elements pertaining to the CI characteristics, including its participants, facilitation, and sustainability efforts. The following questions are typically addressed: *What are the intervention’s objectives, content, and planning?* (D1) *Which team(s) and participants are involved and what characterizes them?* *How is facilitation planned and who will facilitate?* (D2) *What are the sustainability strategies and activities?* (D3). This category also comprises planned and anticipated support outside the training sessions and sustainability activities.

Detailed descriptions of the interventions’ objectives, practicalities (e.g., content, planning, resourcing), as well as their theoretical rationale and related strategies and tactics are imperative for scientific meta-analysis and the replicability of studies (Mohler et al., 2010; Golub & Fontanarosa, 2015; Cheng et al., 2016). Additionally, successful intervention implementation is contingent upon facilitation, scaffolding its roll-out and supporting participants’ engagement and learning. Likewise, the involved team(s) characteristics are essential, such as task types, the differentiation of authority across the team(s), and their stability over time (Hollenbeck et al., 2012; Wildman et al., 2012). Thus, effective reporting comprises detailed descriptions of strategies, tactics, and processes used by an individual (i.e., facilitator) to help others improve their knowledge, skills, or attitudes and thereby improve the intervention’s likelihood of success, including specifications of the facilitator’s qualifications and subject matter expertise (Kitson et al., 1998; Salas et al., 2002). Furthermore, contextual information can be relevant since, for example, small facilities can lack implementation strength, which has to be addressed and anticipated (e.g., with more appropriate facilitation) (Kitson et al., 2008).

Various frameworks also emphasize the relevance of describing sustainment, or maintenance activities, since team training effects at various levels (i.e., reactions, learning, transfer, and outcome) often diminish over time (Arthur et al., 1998; Wierenga et al., 2014). Reports on TS implementation note a decline in training effects within six months to a year (Forse et al., 2011; Thomas & Galla, 2013). Strategies such as regularly planned competency refresher training can help mitigate such declines (Sonesh et al., 2015). Further, concurrent workplace-based coaching and mentoring of staff, as well as post-training sustainability activities, can be applied (Morey et al., 2002; Marshall & Manus, 2007).

Strategies and activities focusing on sustaining the long-term effects of healthcare team training implementation have been suggested to be an underexplored part of implementation and research (Lee et al., 2017). Since “team training is not a one-day or single-session event” (Gillespie et al., 2010, p. 655), the sustainment phase of TS and similar interventions is imperative (AHRQ, 2014). However, only 15 of the 27 articles reviewed here reported such sustainment activities.

III – Execution and deployment (E1-3)

The third category provides components and elements that facilitate reporting effectively and objectively on what occurred during the intervention implementation, including related educational and pedagogical perspectives pertaining to training activities. We thus address the following questions: *What and how much was done—planned and unplanned? What were the deviations from the planned implementation (and why)?* (E1) *How was it done and received?* (E2) *Who delivered the intervention?* (E3)

Our difficulties with rating the included studies’ reports of elements E2 and E3 might have been due to the fact that we selected studies in which a TS master trainer was present. TS master trainers follow a (minimum of) two-and-a-half-day standardized training session, and their roles and tasks are well described in the TS documentation, which possibly kept the researchers from providing extensive reports on this issue. Additionally, we realized that the checklist’s E3 (‘Faculty’) component overlapped somewhat with the D2 elements (‘Facilitation’). While comprehensive multidisciplinary team training curricula, such as TS, are on the rise, healthcare organizations are increasingly applying customized ‘self-made’ approaches, using a variety of trainers, coaches, and other faculty members (Gross et al., 2019). We suggest that, due to the essential role of these individuals in teaching, coaching, and assisting others, the reports should describe their roles in detail, using both elements.

IV – Mechanisms of impact (M1)

The fourth category of the ReCoMuTe checklist provides a set of elements to assist in detailed reporting on impact evaluation. These help to assess possible causal mechanisms and relationships between actual training and facilitatory activities vis-à-vis effects, or lack thereof. Such evaluations of interventions and detailed analyses of either facilitating or impeding mediators, unexpected pathways, and consequences help to answer: *Why and how did the change (not) happen the way it happened?*

Explaining the intervention’s mechanisms of impact can encompass the dynamics that interrelate or overlap across the checklist’s categories and their components. Reports that observed or narrated facilitating and impeding factors as part of a process evaluation provide essential information for, for example, further successful replication. Additionally, an initial inventory of determinants serves authors with a practicable instrument for more complex assessments of what happened. Explicating in detail the ‘why’ of the observed and measured effects, as well as the often unexpected and tacit dynamics imparted by the implementation efforts, requires authors’ reports to be based on a convening of viewpoints, including change management and implementation science. The table resulting from our thematic analysis of the 27 included publications on TS implementation provides an exemplary multi-study overview of determinants (Table 4).

Moderators

Facilitating and impeding factors are moderators that intentionally or unintentionally regulate or determine an implementation process and/or its outcomes (Fleuren et al., 2004). Such factors can be described by their characteristics, which can be classified into five categories: (1) socio-political context (e.g., other interfering interventions; regulations; professionalism); (2) organization (e.g., culture; leadership; resources); (3) facilitator/ implementer (e.g., skills; background; profession); (4) intervention program (e.g., timing; content; complexity); and (5) participants (e.g., participation; attitude; profession; prior experiences) (Wierenga et al., 2012; Wierenga et al., 2013). Our thematic coding of the included studies as ‘facilitating’ or ‘impeding’ revealed that various, mostly facilitating, factors were reported (Table 4).

Table 4. Factors reported as (a) facilitating or (b) impeding TeamSTEPPS implementation in the 27 publications

Categories ¹	Subcategories	Facilitating factors	Impeding factors
1. Socio-political context		<ul style="list-style-type: none"> · Hospital-wide central meetings to share best practices [12]² 	<ul style="list-style-type: none"> · Other interfering safety interventions [16]
2. Organization	Culture	<ul style="list-style-type: none"> · Open communication and mutual respect [5] 	<ul style="list-style-type: none"> · Work environment did not support training, learning, or transfer [5]
	Leadership	<ul style="list-style-type: none"> · Senior physician leadership met with other physicians [15] · Support by leadership [1, 2, 4, 5, 6, 7, 8, 12, 13] · Change team recognized a need for change [2] · Change ownership was taken up and driven by local change team [8] 	<ul style="list-style-type: none"> · Difficult to convince leadership [13]
	Resources	<ul style="list-style-type: none"> · Plenty of time [10] · Infrastructure was already present [4, 27] · Overtime training hours were budgeted [12] 	<ul style="list-style-type: none"> · Lack of resources (sufficient finances, training time) [4, 8, 10, 23, 26]
	Turnover	<ul style="list-style-type: none"> · Low staff turnover [4, 12] 	<ul style="list-style-type: none"> · High staff turnover [10]
	Job position of the implementer	<ul style="list-style-type: none"> · Trainers from different disciplines [6, 8, 13, 19] · Physicians trained by other physicians [13] 	
	Collaboration/interaction	<ul style="list-style-type: none"> · Allowing participants to contribute their own ideas/opinions [7, 8, 14] · Creation of video vignettes and scenarios achieved a high level of staff resonance and buy-in [6] 	
3. Facilitator	Degree of rewards	<ul style="list-style-type: none"> · Encourage use of learned skills by handing out small aids [5, 27] 	
	Preparation	<ul style="list-style-type: none"> · Participating in questionnaires about patient safety [10] · Providing information before training [5] 	
4. Intervention program	Sustainment	<ul style="list-style-type: none"> · Encouraging use of learned skills by team leaders [5] · Additional learning opportunities (to classroom training) [10] · Repetition of training during iterative simulations and pre-shift briefs [17] · Coaching of staff in desired teamwork behavior [1] · Success stories narrated by staff ('storytelling') [18] 	
	Complexity		<ul style="list-style-type: none"> · Lack of guidance on training deployment [2]
	Timing of intervention activities	<ul style="list-style-type: none"> · Free from clinical duties to enable participation [1, 10, 19, 25] · Trainings scheduled at convenient times, minimizing work interruptions [1, 14] · Debriefings during lunch [12] · Sessions scheduled during department meetings to improve attendance [15] 	<ul style="list-style-type: none"> · Difficult for physicians to attend training while on duty [25]
	'Fit' of intervention	<ul style="list-style-type: none"> · Teaching principles specifically targeted at adult learners [8] · Linking concepts of training to practice/clinical experiences [9] 	
5. Participants	Participation	<ul style="list-style-type: none"> · Multidisciplinary participation [5, 8, 13, 24] · Physician participation [13] · Voluntary enrolment [1, 27] 	<ul style="list-style-type: none"> · Lack of multidisciplinary participation [15, 16, 21] · Lack of physician participation [9, 12, 26] · Physicians had to be trained with an abbreviated program [4, 12]

¹ The categories in this table are based on Wierenga et al. (2013).

² The numbers in square brackets correspond to the numbers of the articles in Table 3.

Discussion

In the absence of a preexisting standard, we embarked on developing a checklist for reporting complex interventions, comprising multidisciplinary healthcare TT. Our effort was informed by current frameworks for scientific reports on the design, development, implementation, and evaluation of CTs in healthcare (Möhler et al., 2013; Möhler et al., 2015; Moore et al., 2015), as well as approaches to evaluating team training implementation (Weaver et al., 2010; Zhu et al., 2015) and worksite health promotion programs (Wierenga et al., 2014). Moreover, by consulting extant key literature on related topics and studying other frameworks and guidelines relevant to the domain or quality improvement and teamwork in healthcare, we synthesized the retrieved data following an iterative method that resulted in the new ReCoMuTe checklist.

The final checklist does not directly provide structures derived from general theories, such as implementation (Damschroder et al., 2009) or teamwork (e.g., the input-mediator-output framework; Ilgen et al., 2005). Additionally, we do not include distinct levels for training outcome assessment (Kirkpatrick, 1994). Although only a few of the publications in our review used this approach for assessing the effectiveness of their interventions, we contend that such a comprehensive method for training evaluation should be regarded as a standard procedure (Gross et al., 2019).

During the checklist's development, we did not find applicable approaches for categorizing team training into sub-types or domains. Despite some attempts, this field is presumably too young for such distinctions (Husebo & Akerjordet, 2016). However, the need for improved classifications of teamwork development interventions was increased by recent reports discussing the wide variety of interpretations and jargon, which further impedes optimal reporting in publications (Gross et al., 2019). Therefore, we concur with West and Lyubovnikova (2013) that the 'team' concept must be viewed with some caution, since teams in healthcare can vary widely in, for example, skills, culture, task-focus, and temporal stability or 'fluidity'.

However, scientific research publications primarily focus on aligning with the author guidelines presented by the journals. The ReCoMuTe checklist provides a structured overview of the relevant elements when reporting CTTIs and their implementation. Therefore, while developing our checklist we chose not to include standard items and sections, such as those provided by the IMRaD structure (i.e., introduction, methods, results, and discussion) (Moss & Thompson, 1999). Furthermore, sections of the publications containing, for example, a study's rationale, ethical considerations, methodological approaches, and statistical results, were viewed as outside our checklist's scope; more appropriate (reporting) frameworks exist for them. We chose to develop a new checklist over alternatives such as amendments to or extensions of existing frameworks (e.g., Golub & Fontanarosa, 2015; Cheng et al., 2016). The rationale for this lies in our primary focus on reporting quality (for research accumulation purposes). In our view, extant frameworks provide excellent guidance on other essential scientific publication elements. Hence, we would encourage the combined use of ReCoMuTe and other frameworks and guidelines.

Validity

Using the ReCoMuTe checklist on selected reports on the implementation of a standardized TS curriculum affirmed our checklist's applicability and face validity as an overview of relevant elements when reporting. Assessing the 27 included studies with the checklist did not reveal new items; however, reviewing the included studies provided some additional insights regarding their published reports.

While substantiating previous accounts (Husebo & Akerjordet, 2016), we found, not surprisingly, that the majority of studies originated from the U.S., where TS was originally developed. Furthermore, acute and academic settings seem to be popular settings for TS implementation. Possibly, this is due to the natural discourse of healthcare team training science and practice. In the last two decades, an important knowledge base has been sprouting from high-risk industries (e.g., aviation) to similarly risk-prone healthcare settings, in particular ORs and ERs (Gross et al., 2019; Table 2). Aptly, team training practices in other industries show an evolution that comprises several generations of experimenting and developing before reaching a wide consensus on best practices (Helmreich et al., 1999). Moreover, the initial TS version (v1.0) was primarily aimed at such settings. TS was designed to provide a tailored-to-context set of approaches and building blocks for various settings and, recently, AHRQ has provided new versions adapted for non-OR settings (e.g., long-term care or office-based clinics). While our results corroborate previous references to TS interventions as being primarily used in university hospitals (Ward et al., 2015), recent research on adverse incidents suggests that TS and similar interventions are relevant to regional hospitals as well (Langelaan et al., 2017).

Rating the sample of publications in our review using the ReCoMuTe checklist results in a low reporting score. Apparently, despite TS's highly standardized features, there is no consensus on the 'what-and-how' when reporting evaluations of TS or similar programs. Effective reporting might not be within the scope of those who report (Möhler et al., 2010; Golub & Fontanarosa, 2015). Arguably, this is due to the relative novelty and complexity of the field for these types of healthcare team training interventions and the young age of one of the field's pioneering programs (i.e., TS curriculum) (Clancy & Tornberg, 2007). Promisingly, however, the authors reported elements regarding contextualization, facilitation, and fidelity/adaptability in relative detail (see, Table 1, components C3, D2, and E1).

Medical engagement and leadership

The importance of physician engagement in TS implementation was noted in several studies in our sample. While the multidisciplinary character of TS is reported as an important determinant for implementation success in six studies (Mayer et al., 2011; Brodsky et al., 2013; Thomas & Gala, 2013; Spiva et al., 2014; Amaya-Anyas et al., 2015; Sonesh et al., 2015; Lisbon et al., 2016), it is also explicitly mentioned that physician participation had a similar impact (Jones et al., 2013a; Sheppard et al., 2013; Thomas & Gala, 2013; Wong et al., 2016). Difficulties in scheduling around on-call duties resulted in poor physician participation and a need to adjust programs into abbreviated, physician-only versions (Forse et al., 2011; Sheppard et al., 2013). Interestingly, one study mentioned the beneficial effects of peer-to-peer training among physicians (Thomas & Galla, 2013).

It has been acknowledged for some time that physician competence in effective engagement in quality and safety improvement requires a more integrated educational system (Aron & Headrick, 2002). Also, a "lack of training focus to address hierarchical differences and incivility" has been mentioned as an

impeding factor for TS implementation (Clapper & Ng, 2013, p. 287). Soklaridis and colleagues (2007) found that a barrier to team collaboration was limited formalized education on interprofessional collaboration for physicians, and even fewer training opportunities for other healthcare professionals. As several studies reported the lack of physician involvement as a possible reason for underperforming implementation success (Jones et al., 2013a; Sheppard et al., 2013; Wong et al., 2016), it seems important to address this issue.

Clapper and Ng recommend a preparatory adjustment of existing team culture and improving physicians' leadership skills when implementing TS (Clapper and Ng, 2013). Interestingly, the recently emerged concept of medical leadership has not only gained significant attention inside the medical profession and beyond, but also directs physicians' education and training particularly toward embracing behaviors and non-technical (leadership) skills in teamwork (Dath et al., 2015; Keijser, 2019). Healthcare teamwork and medical leadership have been suggested to be inextricably connected as bilaterally scaffolding principles (Husebo & Akerjordet, 2016; Salas et al., 2018). It is well-reported that physicians and their historically sovereign professional status have a significant impact on psychological safety and, consequentially, on engagement in quality improvement work, such as CTTIs (Nemhard & Edmondson, 2006). Contrastingly, new insights into medical education suggest more collective practice-based learning, such as in CTTIs, as essential strategies in educating physicians (Cruess et al., 2018). However, because of insufficient reporting in publications describing multidisciplinary teamwork training, including physicians, it often remains unclear if and how the challenges regarding physicians' involvement and their leadership can be addressed.

Practical Implications

The ReCoMuTe checklist could serve to help upgrade the comprehensiveness of reporting intervention studies. These include investigations of the efficacy of TT as an intervention for team performance improvement and studies using TT as an investigative approach to scrutinize specific topics, such as quality improvement, medical leadership development, interprofessional learning, or organizational transformation. Establishing and assessing research evidence supporting CTTIs might not always be the primary objective of field implementors and their teams (Ward et al., 2017). Yet, we hope our checklist will also provide a practical instrument for these groups.

The ReCoMuTe checklist importantly aims to assist authors of papers involving TT and to also be instrumental for editors and reviewers when assessing the suitability of submitted work to enhance their readership's access to detailed accounts of experiences and learned lessons. Although we set out to facilitate the work of designers, implementers, and evaluators of complex, multidisciplinary TT interventions, we believe the current checklist also serves as a guiding instrument for less complex types of multidisciplinary TT programs. Arguably, within the longitudinal perspective of a few weeks to several years (i.e., typical range for implementing team training programs), any implementation effort comprising multidisciplinary and organizational dynamics can spiral into complexity at some point (Husebo & Akerjordet, 2016). Moreover, context as well as interventions tend to change during the phases of implementation and sustainability activities, which is typically the period that is covered in scientific reports. Hence, there is a need to describe such evolutions and their related dynamics (e.g., unexpected changes impacting participants' engagement), which can be aided by the ReCoMuTe checklist.

Ultimately, well-constructed publications can feed into meta-analyses, substantiating the well-informed design and delivery of CTTIs. However, RCTs have only entered the field of healthcare team training recently (Panella et al., 2012; Husebo & Akerjordet, 2016; Strasser et al., 2018; Dinius et al., 2019). Since RCTs represent the 'Gold standard' in evidence-based practice, this possibly heralds a certain timeliness for the ReCoMuTe checklist (Craig et al., 2008; Bhattacharyya et al., 2009). We hope that our checklist will help to motivate those engaging in meta-analytic endeavors of meticulous comparative studies and to embolden them in their demanding review processes to be rigorous in judging a report's eligibility, and maybe even disregard those that have 'stripped away' information on essential contextual matters (Booth et al., 2019; Greenhalgh, 2012).

Due to journals' wordcount limits, authors increasingly report details through an extra publication, (e.g., a study protocol) or a website (Möhler et al., 2015). Furthermore, since almost no clinicians and only a few researchers have time to personally contact authors requesting additional information, editors should encourage their submitting authors to allocate locations beyond their primary publication (e.g., online supplementary documents) (Hoffmann et al., 2014).

We emphasize, though, that the ReCoMuTe checklist is not a 'recipe,' nor a comprehensive guideline. It is not a complete list of 'must-haves' to direct the design or implementation of CTTIs. Moreover, the scientific relevance of some of the elements to a CTTI's effectiveness is neither established nor disputed (e.g., team composition, work environment, training strategy) (Hughes et al., 2016). Instead, it is a starting point for those endeavoring to research CTTIs and their implementation. The checklist comprises a structured overview of known elements, and only future research can reveal the causal relationship for most of them with impact mechanisms and effectiveness. Therefore, some caution in using the new ReCoMuTe checklist in further work is warranted, for which we provide the following considerations.

Limitations

We have several considerations regarding the development and testing of the ReCoMuTe checklist. First, we used a non-exhaustive and pragmatic approach, which was not based on extensive consensus building (e.g., expert group consultation); it addressed our own need for adequate reporting, suitable for the purpose of knowledge accumulation, given the abundance of existing reporting standards (Simera et al., 2010; Pinnock et al., 2015; Bragge et al., 2017). An imaginable selection bias, instilled by a possibly incomplete list of sources, was mitigated by our stepped development of the iteratively adapted consecutive versions of the checklist. The combination of extensive scholarly and practical experience in TT, organizational change, medical leadership, and healthcare transformation within our group, as well as including an initial validity test, also mitigated bias. Follow-up work could comprise expert surveys and consensus meetings, including Delphic methods, which could build on the checklist presented by the present study (Möhler et al., 2010b; Hoffmann et al., 2014). Since the TT field is certainly expanding rapidly and in various dimensions, the ReCoMuTe checklist will need, like most reporting guidelines, revision or refinement. Above all, we hope that we have stepped up the discipline used by authors and editors of journals for all the reporting criteria voiced in this paper.

Second, using one distinct type of CTTI (i.e., TeamSTEPPS) for face validation may have instilled a bias. Conceivably, using a wider spectrum of CTTI approaches for validity testing could have revealed elements that were not part of this study. Although TS's unequivocal tailorable characteristics offer

significant variability in its application in practice, using clear reporting criteria (i.e., the ReCoMuTe checklist) might help prevent publications on TS, or similar programs, from becoming a heterogeneous 'hodge-podge,' and thus from reducing their potential (Petticrew et al., 2013).

Third, the present study's focus on intervention, implementation, and process evaluation, included elements of the 'partial system perspective.' This enabled the capturing of information on the interplay between environmental and organizational aspects, such as processes, cultures, and values, and the interventions themselves. However, meso and macro system perspectives (e.g., economic, social, and political factors; regulatory matters; policies) must also be considered when consolidating results from various sources to produce transferable findings (Booth et al., 2019). Healthcare systems' reform and transformation endeavors should, ultimately, be aligned with the level of the teams and their performance, since it is at this level that healthcare is created and delivered (Keijser & Martin, 2019).

We believe these considerations will encourage eventual further testing of the ReCoMuTe checklist. The efforts should preferably include other settings than academic and acute types, since the significance of improving patient safety and healthcare quality certainly extends beyond these settings as well. Further work could entail publication of an 'explanation and elaboration paper,' providing exemplary research questions, discussing methodologies, advising on data collection, and discussing exemplary publications. Parallel to this, based on our current work, it is our opinion that the TT domain needs a more established operational definition of teamwork and training approaches, including a glossary of concepts, terminology, and instruments that would aid future standardization and subsequent reporting.

In sum, we believe the status quo of TT reporting is in its early stages of growing toward maturity. Suggested future work in this area should exceed the scope and depth of the present study and must contribute to a more comprehensive type of reporting guideline or standard. Until then, the (current version of the) ReCoMuTe checklist provides a structured set of elements that has been proven to be useful for analyzing reports on TS interventions, so that further studies and reports on those interventions could, eventually, save more lives.

Conclusions

This paper originates from the authors' impression of inadequate reporting of CTTIs, and an increasing call for standards to enhance completeness, reproducibility, and comparability of TT (Mohler et al., 2015; Gross et al., 2019). Incomplete descriptions of interventions in publications on CTTI hinders an effective use of resources invested in healthcare TT and the evaluation of its implementation. As expected, we found in this systematic literature-review study that the current quality of reporting CTTIs is remarkably poor. Hence, better-quality reporting on the 'what, how, who, when, and why' of CTTIs is urgently needed. Due to this inadequate reporting, we cannot point out causal relationships and dynamics relating to the failure or success of TT implementation. The complexity of the phenomenon under study, as well as the multidisciplinary perspectives taken, fuel the need for adequate evaluations or reports on TT interventions.

We present in this paper the new ReCoMuTe checklist which can be used to step up the quality and completeness of reporting on multidisciplinary TT in complex intervention formats to enable, ultimately, better replication of these interventions, fostering a better understanding of the necessary (professional) behaviors and other (e.g., contextual) factors involved. The ReCoMuTe checklist can be used by: authors to structure their reports on the interventions; editors and reviewers to evaluate the completeness of the offered descriptions; and readers or practitioners interested in an effective design and implementation of TT. The checklist may also support the planning of purposeful research agendas, while also contributing to wise spending of resources in the current generation of varied CTTIs and related research activities (Hoffmann, 2013; Möhler et al., 2013). We tested the ReCoMuTe checklist using publications reporting on the implementation of TS.

We found that TS as described in studies we included has been primarily implemented in high-demand settings in U.S.-based university hospitals, and that training effects were mostly researched using pre-post training and longitudinal research designs. Also, these publications generally reported the importance of multidisciplinary participation and physician engagement. Not surprisingly, most of the TS publications included in our systematic analysis did not report a majority of the elements of the ReCoMuTe checklist.

Our study affirms that, given the extant reporting levels, more adequate reporting is needed in this area to foster more practical use and replication. We also note a 'tailored-standardized paradox'. The use of a comprehensive type of CTTI, such as TS, which is explicitly designed to be implemented in a tailored manner, based on local, or contextual, needs and demands, imparts a variability in implementation modes, which makes comparability between single implementation reports difficult. Considering these non-linear characteristics of implementing CTTI programs in healthcare, further work on standardization, such as for reporting, can help scholars and practitioners find a balance between situational adaptability and standardization (Chen et al., 2019). We contend that, like the evolutionary development of healthcare quality improvement science and TT in other industries, the emerging science of complex and multidisciplinary TT interventions is gradually becoming 'institutionalized' but it is not sufficiently 'professionalized' yet (Audet et al., 2005). However, TT and related (e.g., leadership) developments are becoming integral elements of audits, professionalization, and educational schemes in healthcare, which is a similar process to how TT evolved in the field of aviation (EASA, 2018; Rosenbaum, 2019).

Our study also stresses physicians' crucial role and position in CTTIs, in line with their relevance to effective system reform and safety culture (Nembhard & Edmondson, 2006; Porter & Tesiberg, 2007). Physician involvement, or lack thereof, in CTTIs, has a significant influence on the failure or success of implementation efforts. This indicates a need to ensure effective investments in assisting the medical profession and its members, for instance through enhancing their leadership competency development and a reciprocal adaptability between various professional identities (Bååthe et al., 2013; Jones et al., 2013b; Sheppard et al., 2013; Wong et al., 2016; Keijser, 2019).

Stepping back from the present study to a more macro-level perspective, we suggest decision makers to consider a strategic, synchronous deployment of CTTIs across their organizations' departments and teams. Since healthcare professionals tend to work within multiple clinical teams, more strategically

organized rollout of often highly resourced CTTIs initiatives could mitigate a potential dilution of TT's beneficial effects at individual and local team levels. However, to be effective, top-down support, endorsement and governance require adequate multidisciplinary organizational knowledge and experience at the executive and managerial level (Salas et al., 2002; Lyubovnikova et al., 2018; Salas et al., 2018). More sophisticated healthcare management will therefore contribute to more collective (and academically sound) sharing of experiences, expertise (e.g., trainers), and other resources. Ultimately, inclusive multi-level investments in optimal collective teamwork and leadership is required for system reform, transformation, and substantive improvement (De Brún et al., 2019). We hope the ReCoMuTe checklist will provide a beneficial contribution to this end.

The ReCoMuTe checklist was registered in the EQUATOR Network on July 15th, 2019 (See, <https://www.equator-network.org/reporting-guidelines/>).

Abbreviations

- AHRQ, Agency for Healthcare Research and Quality
- CI, Complex Intervention
- CTTI, Complex Team Training Intervention
- ED, Emergency Department
- ICU, Intensive Care Unit
- L&D, Labor and Delivery
- NICU, Neonatal Intensive Care Unit
- OR, Operation Room
- ReCoMuTe, Reporting Complex Multidisciplinary Healthcare Teamwork Training
- TS, TeamSTEPPS
- TT, Team Training

References

1. (AHRQ) AfHRAQ. TeamSTEPPS 2.0 Course Management Guide 2014. <https://www.ahrq.gov/teamstepps/instructor/essentials/coursemgmt.html>. Accessed 26 Mar 2020.
2. (EASA) EASA. Commission Regulation (EU) No 965/2012 on air operations in combination with AMC1 ORO. 2018;115.
3. Amaya-Añas AC, Idarraga D, Giraldo V, Gómez LM. Effectiveness of a program for improving teamwork in operating rooms. *Revista Colombiana de Anestesiología*. 2015;43(1):68–75.
4. Andreatta PB. A typology for healthcare teams. *Health Care Manage Rev*. 2010;35(4):345–54.
5. Aron DC, Headrick L. Educating physicians prepared to improve care and safety is no accident: It requires a systematic approach. *Qual Saf Health Care*. 2002;11(2):168–73.
6. Arora S, Hull L, Fitzpatrick M, Sevdalis N, Birnbach DJ. Crisis management on surgical wards: A simulation-based approach to enhancing technical, teamwork, and patient interaction skills. *Ann Surg*. 2015;261(5):888–93.
7. Arthur W Jr, Bennett W Jr, Stanush PL, McNelly TL. Factors that influence skill decay and retention: A quantitative review and analysis. *Hum Perform*. 1998;11(1):57–101.
8. Audet A-MJ, Doty MM, Shamasdin J, Schoenbaum SC. Measure, learn, and improve: Physicians' involvement in quality improvement. *Health Aff*. 2005;24(3):843–53.
9. Bååthe F, Erik Norbäck L. Engaging physicians in organisational improvement work. *J Health Organ Manag*. 2013;27(4):479–97.
10. Beitlich P. TeamSTEPPS implementation in the LD/NICU settings. *Nurs Manag*. 2015;46(6):15–8.
11. Berwick DM. Developing and testing changes in delivery of care. *Ann Intern Med*. 1998;128(8):651–6.
12. Bhattacharyya O, Reeves S, Zwarenstein M. What is implementation research? Rationale, concepts, and practices. *Res Soc Work Pract*. 2009;19(5):491–502.
13. Booth A, Moore G, Flemming K, Garside R, Rollins N, Tunçalp Ö, et al. Taking account of context in systematic reviews and guidelines considering a complexity perspective. *BMJ Glob Health*. 2019;4(S1):e000840.
14. Bragge P, Grimshaw JM, Lokker C, Colquhoun H. AIMD-a validated, simplified framework of interventions to promote and integrate evidence into health practices, systems, and policies. *BMC Med Res Method*. 2017;17(1):38.
15. Brewster L, Aveling E-L, Martin G, Tarrant C, Dixon-Woods M. What to expect when you're evaluating healthcare improvement: a concordat approach to managing collaboration and uncomfortable realities. *BMJ Qual Saf*. 2015;24(5):318–24.
16. Brodsky D, Gupta M, Quinn M, Smallcomb J, Mao W, Koyama N, et al. Building collaborative teams in neonatal intensive care. *BMJ Qual Saf*. 2013;22(5):374–82.
17. Buljac-Samardzic M, Dekker-van Doorn CM, van Wijngaarden JD, van Wijk KP. Interventions to improve team effectiveness: A systematic review. *Health Pol*. 2010;94(3):183–95.

18. Campbell M, Fitzpatrick R, Haines A, Kinmouth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ*. 2000;321:694–6.
19. Capella J, Smith S, Philip A, Putnam T, Gilbert C, Fry W, et al. Teamwork improves the clinical care of trauma patients. *J Surg Educ*. 2010;67(6):439–43.
20. Chen AS, Yau B, Revere L, Swails J. Implementation, evaluation, and outcome of TeamSTEPPS in interprofessional education: A scoping review. *J Interprof Care*. 2019;33(6):795–804.
21. Cheng A, Kessler D, Mackinnon R, Chang TP, Nadkarni VM, Hunt EA, et al. Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. *Adv Simul*. 2016;1(1):238–48.
22. Clancy CM, Tornberg DN. TeamSTEPPS: Assuring optimal teamwork in clinical settings. *Am J Med Qual*. 2007;22(3):214–7.
23. Clapper TC, Kong M. TeamSTEPPS™. The patient safety tool that needs to be implemented. *Clinical Simulation in Nursing*. 2012;8(8):e367-e73.
24. Clapper TC, Ng GM. Why your TeamSTEPPS™ program may not be working. *Clin Simul Nurs*. 2013;9(8):e287-e92.
25. Clay-Williams R, Braithwaite J. Reframing implementation as an organisational behaviour problem: Inside a teamwork improvement intervention. *J Health Organ Manag*. 2015;29(6):670–83.
26. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *JAMA*. 2011;306(9):978–88.
27. Cook DA, Levinson AJ, Garside S. Method and reporting quality in health professions education research: A systematic review. *Med Educ*. 2011;45(3):227–38.
28. Council MR. A framework for the development and evaluation of RCTs for complex interventions to improve health. 2000. <https://mrc.ukri.org/documents/pdf/rcts-for-complex-interventions-to-improve-health/>. Accessed 26 Mar 2020.
29. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*. 2008;337:1655.
30. Cruess RL, Cruess SR, Steinert Y. Medicine as a community of practice: Implications for medical education. *Acad Med*. 2018;93(2):185–91.
31. Damschroder LJ, Aron DC, Keith RE, Kirsch SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*. 2009;4(1):50.
32. Dath D, Chan MK, Abbott C. CanMEDS 2015: from manager to leader. Ottawa: Royal College of Physicians and Surgeons of Canada; 2015.
33. De Brún A, O'Donovan R, McAuliffe E. Interventions to develop collectivistic leadership in healthcare settings: a systematic review. *BMC Health Serv Res*. 2019;19(1):72.
34. Deering S, Rosen MA, Ludi V, Munroe ML, Pocrnich A, Laky C, et al. On the front lines of patient safety: implementation and evaluation of team training in Iraq. *Jt Comm J Qual Pat Saf*. 2011;37(8):350–6.
35. Deneckere S, Euwema M, Van Herck P, Lodewijckx C, Panella M, Sermeus W, et al. Care pathways lead to better teamwork: Results of a systematic review. *Soc Sci Med*. 2012;75(2):264–8.
36. Dinius J, Hammer A, Manser T, Bergelt C, Kriston L, Körner M. Piloting and evaluating feasibility of a training program to improve patient safety for inter-professional inpatient care teams—study protocol of a cluster randomized controlled trial. *Trials*. 2019;20(1):386.
37. Dixon-Woods MAR, Bosk CL, Aveling EL, Goeschel CA, Pronovost PJ. Explaining Michigan: Developing an ex post theory of a quality improvement program. *Milbank Q*. 2011;89(2):167–205.
38. Fischer MM, Tubb CC, Brennan JA, Soderdahl DW, Johnson AE. Implementation of TeamSTEPPS at a Level-1 military trauma center: The San Antonio Military Medical Center experience. *US Army Medical Department*; 2015(Oct.-Dec.):75 – 9.
39. Fleuren M, Wiefferink K, Paulussen T. Determinants of innovation within health care organizations: Literature review and Delphi study. *Int J Qual Health Care*. 2004;16(2):107–23.
40. Forse RA, Bramble J, McQuillan R. Team training can improve operating room performance. *Surgery*. 2011;150(4):771–8.
41. Foy R, Ovreteit J, Shekelle PG, Pronovost PJ, Taylor SL, Dy S, et al. The role of theory in research to develop and evaluate the implementation of patient safety practices. *BMJ Qual Saf*. 2011;20(5):453–9.
42. Gaston T, Short N, Ralvea C, Casterline G. Promoting Patient Safety: Results of a TeamSTEPPS™ Initiative. *J Nurs Admin*. 2016;46(4):201–7.
43. Gillespie BM, Chaboyer W, Murray P. Enhancing communication in surgery through team training interventions: a systematic literature review. *AORN J*. 2010;92(6):642–57.
44. Glasgow BM, Harden SM, Gaglio B, Rabin BA, Smith ML, Porter PC, et al. RE-AIM planning and evaluation framework: Adapting to new science and practice with a twenty-year review. *Front Public Health*. 2019;7:64.
45. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Pub Health*. 1999;89(9):1322–7.
46. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267–76.
47. Golub RM, Fontanarosa PB. Researchers, readers, and reporting guidelines: writing between the lines. *JAMA*. 2015;313(16):1625–6.
48. Goodman D, Ogrinc G, Davies L, Baker GR, Barnsteiner J, Foster TC, et al. Explanation and elaboration of the SQUIRE (Standards for Quality Improvement Reporting Excellence) Guidelines, V. 2.0: examples of SQUIRE elements in the healthcare improvement literature. *BMJ Qual Saf*. 2016;25(12):986–92.
49. Gough D, Thomas J, Oliver S. Clarifying differences between review designs and methods. *Syst Rev*. 2012;1(1):28.
50. Greenhalgh T. Outside the Box: Why are Cochrane reviews so boring? *Brit J Gen Pract*. 2012;62(600):371.

51. Gross B, Rusin L, Kiesewetter J, Zottman JM, Fischer MR, Prückner S, et al. Crew resource management training in healthcare: A systematic review of intervention design, training conditions and evaluation. *BJM Open*. 2019;9(2).
52. Gupta RT, Sexton JB, Milne J, Frush DP. Practice and quality improvement: successful implementation of TeamSTEPPS tools into an academic interventional ultrasound practice. *Am J Roent*. 2015;204(1):105–10.
53. Haerens MH, Kox M, Noe PM, Van Der Hoeven JG, Pickkers P. Crew resource management in the trauma room: a prospective 3-year cohort study. *Eur J Emerg Med*. 2018;25(4):281–7.
54. Hawe P, Shiell A, Riley T. Theorizing interventions as events in systems. *Am J Comm Psychol*. 2009;43(3–4):267–76.
55. Helmreich RL, Merritt AC, Wilhelm JA. The evolution of crew resource management training in commercial aviation. *Int J Aviat Psychol*. 1999;9(1):19–32.
56. Hoffmann TC, Eructi C, Glasziou PP. Poor description of non-pharmacological interventions: Analysis of consecutive sample of randomised trials. *BMJ*. 2013;347.
57. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014;348:g1687.
58. Hollenbeck JR, Beersma B, Schouten ME. Beyond team types and taxonomies: A dimensional scaling conceptualization for team description. *Ac Manag Rev*. 2012;37(1):82–106.
59. Hughes AM, Gregory ME, Joseph DL, Sonesh SC, Marlow SL, Lacerenza CN, et al. Saving lives: A meta-analysis of team training in healthcare. *J Appl Psych*. 2016;101(9):1266–304.
60. Husebø SE, Akerjordet K. Quantitative systematic review of multi-professional teamwork and leadership training to optimize patient outcomes in acute hospital settings. *J Adv Nurs*. 2016;72(12):2980–3000.
61. Ilgen DR, Hollenbeck JR, Johnson M, Jundt D. Teams in organizations: From input-process-output models to IMOI models. *Ann Rev Psych*. 2005;56:517–43.
62. Jones F, Podila P, Powers C. Creating a culture of safety in the emergency department: the value of teamwork training. *J Nurs Adm*. 2013a;43(4):194–200.
63. Jones KJ, Skinner AM, High R, Reiter-Palmon R. A theory-driven, longitudinal evaluation of the impact of team training on safety culture in 24 hospitals. *BMJ Qual Saf*. 2013b;22(5):394–404.
64. Keijser W. Opening Up Medicine: Physicians and Leadership in Times of Transformation. Enschede: University of Twente; 2019.
65. Keijser W, Huq J-L, Reay T. Enacting medical leadership to address wicked problems. *BMJ Leader*. 2020;4:12–7.
66. Keijser W, Martin G. Unlocking medical leadership's potential: a multilevel virtuous circle? *BMJ Leader*. 2020;4:6–11.
67. Kemper PF, van Dyck C, Wagner C, Wouda L, de Bruijne M. Barriers and facilitators for taking action after classroom-based crew resource management training at three ICUs. *Jt Comm J Qual Patient Saf*. 2014;40(7):311–8.
68. King HB, Battles J, Baker DP, et al. TeamSTEPPS™: team strategies and tools to enhance performance and patient safety. Rockville: Agency for Healthcare Research and Quality; 2008.
69. Kirkpatrick DL. Evaluating Training Programs: The Four Levels. San Fransisco: Berret-Koehler; 1994.
70. Kitson A, Harvey G, McCormack B. Enabling the implementation of evidence-based practice: A conceptual framework. *BMJ Qual Saf*. 1998;7(3):149–58.
71. Kitson AL, Rycroft-Malone J, Harvey G, McCormack B, Seers K, Titchen A. Evaluating the successful implementation of evidence into practice using the PARIHS framework: theoretical and practical challenges. *Implement Sci*. 2008;3(1):1.
72. Klipfel JM, Carlolan BJ, Brytowski N, Mitchell CA, Gettman MT, Jacobson TM. Patient safety improvement through in situ simulation interdisciplinary team training. *Urol Nurs*. 2014;34(1):39.
73. Kohn LT, Corrigan JM, Donaldson MS. To Err Is human building a safer health system. Washington: National Academies Press; 2000.
74. Koning H, Verver JP, Heuvel J, Bisgaard S, Does RJ. Lean six sigma in healthcare. *J Healthc Qual*. 2006;28(2):4–11.
75. Langelaan M, Broekens MA, de Bruijne MC, de Groot JF, Moesker MJ, Porte PJ, et al. Monitor Zorggerelateerde Schade 2015/2016; dossieronderzoek bij overleden patiënten in Nederlandse ziekenhuizen 2017. www.nivel.nl. Accessed 26 Mar 2020.
76. Langelaan M, de Bruijne MC, Baines RJ, Broekens MA, Hammink K, Schilp J, et al. Monitor Zorggerelateerde Schade 2011/2012: dossieronderzoek in Nederlandse ziekenhuizen. 2013. . Accessed 26 Mar 2020.
77. Lee S-H, Khanuja HS, Blanding RJ, Sedgwick J, Pressimone K, Ficke JR, Jones LC. Sustaining Teamwork Behaviors Through Reinforcement of TeamSTEPPS Principles. *J Patient Saf*. 2017. doi:10.1097/PTS.0000000000000414.
78. Leonard M, Graham S, Bonacum D. The human factor: the critical importance of effective teamwork and communication in providing self care. *BMJ Qual Saf*. 2004;13:85–90.
79. Lisbon D, Allin D, Cleek C, Roop L, Brimacombe M, Downes C, et al. Improved knowledge, attitudes, and behaviors after implementation of TeamSTEPPS training in an academic emergency department: A pilot report. *Am J Med Qual*. 2016;31(1):86–90.
80. Lyubovnikova J, West TH, Dawson JF, West MA. Examining the indirect effects of perceived organizational support for teamwork training on acute health care team productivity and innovation: The role of shared objectives. *Group Organ Manag*. 2018;43(3):382–413.
81. Mahoney JS, Ellis TE, Garland G, Palyo N, Greene PK. Supporting a psychiatric hospital culture of safety. *J Am Psychiatr Nurses Assoc*. 2012;18(5):299–306.
82. Marlow S, Bisbey T, Lacerenza C, Salas E. Performance measures for health care teams: A review. *Small Group Res*. 2018;49(3):306–56.
83. Marlow SL, Hughes AM, Sonesh SC, Gregory ME, Lacerenza CN, Benishek LE, et al. A systematic review of team training in health care: Ten questions. *Jt Comm J Qual Patient Saf*. 2017;43(3):197–204.

84. Marshall DA, Manus DA. A team training program using human factors to enhance patient safety. *AORN J*. 2007;86(6):994–1011.
85. Marshall S, Harrison J, Flanagan B. The teaching of a structured tool improves the clarity and content of interprofessional clinical communication. *BMJ Qual Saf*. 2009;18(2):137–40.
86. May C, Finch T, Mair F, Ballini L, Dowrick C, Eccles M, et al. Understanding the implementation of complex interventions in health care: The normalization process model. *BMC Health Serv Res*. 2007;7(1):148.
87. Mayer CM, Cluff L, Lin W-T, Willis TS, Stafford RE, Williams C, et al. Evaluating efforts to optimize TeamSTEPPS implementation in surgical and pediatric intensive care units. *Jt Comm J Qual Patient Saf*. 2011;37(8):365–74.
88. Moffatt-Bruce SD, Hefner JL, Mekhjian H, McAlearney JS, Latimer T, Ellison C, et al. What is the return on investment for implementation of a crew resource management program at an academic medical center? *Am J Med Qual*. 2017;32(1):5–11.
89. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol*. 2010;63(8):1–37.
90. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7(2):e000217.
91. Möhler R, Bartoszek G, Meyer G. Quality of reporting of complex healthcare interventions and applicability of the CReDECI list-a survey of publications indexed in PubMed. *BMC Med Res Methodol*. 2013;13(1):125.
92. Möhler R, Köpke S, Meyer G. Criteria for Reporting the Development and Evaluation of Complex Interventions in healthcare: Revised guideline (CReDECI 2). *Trials*. 2015;16(1):204.
93. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*. 2015;350:1258.
94. Morey JC, Simon R, Jay GD, Wears RL, Salisbury M, Dukes KA, 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(6):1553–81.
95. Moss FIONA, Thomson R. A new structure for quality improvement reports. *Qual Health Care*. 1999;8:76.
96. Motley CL, Dolansky MA. Five steps to providing effective feedback in the clinical setting: a new approach to promote teamwork and collaboration. *J Nurs Edu*. 2015;54(7):399–403.
97. Nembhard IM, Edmondson AC. Making it safe: The effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *J Organ Behav*. 2006;27(7):941–66.
98. Organization WH. Patient safety curriculum guide: multiprofessional edition. Geneva: World Health Organization; 2011.
99. Øvretveit J. Understanding the conditions for improvement: research to discover which context influences affect improvement success. *BMJ Qual Saf*. 2011;20:i18–23.
100. Panella M, Marchisio S, Brambilla R, Vanhaecht K, Di Stanislao F. A cluster randomized trial to assess the effect of clinical pathways for patients with stroke: Results of the clinical pathways for effective and appropriate care study. *BMC Med*. 2012;10(1):71.
101. Peters VK, Harvey EM, Wright A, Bath J, Freeman D, Collier B. Impact of a TeamSTEPPS trauma nurse academy at a level 1 trauma center. *J Emerg Nurs*. 2018;44(1):19–25.
102. Petticrew M, Anderson L, Elder R, Grimshaw J, Hopkins D, Hahn R, et al. Complex interventions and their implications for systematic reviews: A pragmatic approach. *J Clin Epidemiol*. 2013;66(11):1209–14.
103. Petticrew M, Knai C, Thomas J, Rehfuess E, Noyes J, Gerhardus A, et al. Implications of a complexity perspective for systematic reviews and guideline development in health decision making. *BMJ Glob Health*. 2019;4(S1):e000899.
104. Pfadenhauer LM, Gerhardus A, Mozygemba K, Lysdahl KB, Booth A, Hofmann B, et al. Making sense of complexity in context and implementation: The Context and Implementation of Complex Interventions (CICI) framework. *Implement Sci*. 2017;12(1):21.
105. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, Griffiths CJ, et al. Standards for reporting implementation studies (StaRI) statement. *BMJ*. 2017;356:i6795.
106. Pinnock H, Epiphaniou E, Sheikh A, Griffiths C, Eldridge S, Craig P, et al. Developing standards for reporting implementation studies of complex interventions (StaRI): a systematic review and e-Delphi. *Implement Sci*. 2015;10(1):42.
107. Porter ME, Teisberg EO. How physicians can change the future of health care. *JAMA*. 2007;297(10):1103–11.
108. Reay T, Goodruck E, Hinings B. Institutionalization and professionalization. *The Oxford Handbook of health care management*. Oxford: Oxford University Press; 2016.
109. Rhee AJ, Valentin-Salgado Y, Eshak D, Feldman D, Kischak P, Reich DL, et al. Team training in the perioperative arena: A methodology for implementation and auditing behavior. *Am J Med Qual*. 2017;32(4):369–75.
110. Robichaud P, Saari M, Burnham E, Omar S, Wray RD, Baker R, et al. The value of a quality improvement project in promoting interprofessional collaboration. *J Interprof Care*. 2012;26(2):158–60.
111. Romijn A, Ravelli ACJ, de Bruijne MC, Twisk JWR, Wagner C, de Groot CJM, et al. Effect of a cluster randomized team training intervention on adverse perinatal and maternal outcomes: a stepped wedge study. *BJOG*. 2019;126(7):907–14.
112. Rosen MA, DiazGranados D, Dietz AS, Benishek LE, Thompson D, Pronovost PJ, et al. Teamwork in healthcare: Key discoveries enabling safer, high-quality care. *Am Psychol*. 2018;73(4):433.
113. Rosenbaum L. Divided we fall. *NEJM*. 2019a;380(7):684–8.
114. Rycroft-Malone JO. Evidence-informed practice: From individual to context. *J Nurs Manag*. 2008;16(4):404–8.

115. Salas E, DiazGranados D, Weaver SJ, King H. Does team training work? Principles for health care. *Acad Emerg Med*. 2008;15(11):1002–9.
116. Salas E, Paige JT, Rosen MA. Creating new realities in healthcare: The status of simulation-based training as a patient safety improvement strategy. *BMJ Qual Saf*. 2013;22:449–52.
117. Salas E, Sims DE, Burke CS. Is there a “Big Five” in teamwork? *Small Group Res*. 2005;36:555–99.
118. Salas E, Wilson KA, Burke CS, Bowers CA. Myths about crew resource management training. *Ergon Des*. 2002;10(4):20–4.
119. Salas E, Wilson KA, Murphy CE, King H, Baker D. What crew resource management training will not do for patient safety. unless... *J Patient Saf*. 2007;3(2):62–4.
120. Salas E, Zajac S, Marlow SL. Transforming health care one team at a time: ten observations and the trail ahead. *Group Organ Manag*. 2018;43(3):357–81.
121. Sawyer T, Laubach VA, Hudak J, Yamamura K, Pocrnich A. Improvements in teamwork during neonatal resuscitation after interprofessional TeamSTEPPS training. *Neonatal Netw*. 2013;32(1):26–33.
122. Scotten M, Manos EL, Malicoat A, Paolo AM. Minding the gap: Interprofessional communication during inpatient and post discharge chasm care. *Patient Educ Couns*. 2015;98(7):895–900.
123. Sexton JB, Thomas EJ, Helmreich RL. Error, stress, and teamwork in medicine and aviation: cross sectional surveys. *BMJ*. 2000;320(7237):745–9.
124. Sheppard F, Williams M, Klein VR. TeamSTEPPS and patient safety in healthcare. *J Healthc Risk Manag*. 2013;32(3):5–10.
125. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med*. 2010;8(1):24.
126. Soklaridis S, Oandasan I, Kimpton S. Family health teams: Can health professionals learn to work together? *Can Fam Physician*. 2007;53(7):1198–9.
127. Sonesh SC, Gregory ME, Hughes AM, Feitosa J, Benishek LE, Verhoeven D, et al. Team training in obstetrics: A multi-level evaluation. *Fam Syst Health*. 2015;33(3):250.
128. Spiva L, Robertson B, Delk ML, Patrick S, Kimrey MM, Green B, et al. Effectiveness of team training on fall prevention. *J Nurs Care Qual*. 2014;29(2):164–73.
129. Stead K, Kumar S, Schultz TJ, Tiver S, Pirone CJ, Adams RJ, et al. Teams communicating through STEPPS. *Med J Austral*. 2009;190(11):128.
130. Strasser DC, Falconer JA, Stevens AB, Uomoto JM, Herrin J, Bowen SE, et al. Team training and stroke rehabilitation outcomes: a cluster randomized trial. *Arch Phys Med Rehabil*. 2008;89(1):10–5.
131. Suter E, Arndt J, Arthur N, Parboosingh J, Taylor E, Deutschlander S. Role understanding and effective communication as core competencies for collaborative practice. *J Interprof Care*. 2009;23(1):41–51.
132. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan–do–study–act method to improve quality in healthcare. *BMJ Qual Safety*. 2014;23(4):290–8.
133. Thomas L, Galla C. Building a culture of safety through team training and engagement. *BMJ Qual Safety*. 2013;22(5):425–34.
134. Treadwell J, Binder B, Symes L, Krepper R. Delivering team training to medical home staff to impact perceptions of collaboration. *Profes Case Manag*. 2015;20(2):81–8.
135. Turner P. Implementation of TeamSTEPPS in the emergency department. *Crit Care Nurs Q*. 2012;35(3):208–12.
136. Vanhaecht K, Panella M, Van Zelm R, Sermeus W. An overview on the history and concept of care pathways as complex interventions. *Int J Care Pathw*. 2010;14(3):117–23.
137. Ward MM, Baloh J, Zhu X, Stewart GL. Promoting action on research implementation in health services framework applied to TeamSTEPPS implementation in small rural hospitals. *Health Care Manage Rev*. 2017;42(1):2–13.
138. Ward MM, Zhu X, Lampman M, Stewart GL. TeamSTEPPS implementation in community hospitals: Adherence to recommended training approaches. *Int J Health Care Qual Assur*. 2015;28(3):234–44.
139. Weaver SJ, Dy SM, Rosen MA. Team-training in healthcare: A narrative synthesis of the literature. *BMJ Qual Saf*. 2014;23:359–72.
140. Weaver SJ, Lyons R, DiazGranados D, Rosen MA, Salas E, Oglesby J, et al. The anatomy of health care team training and the state of practice: A critical review. *Acad Med*. 2010;85(11):1746–60.
141. West C, Landry K, Graham A, Graham L, Cianciolo AT, Kalet A, et al. Conceptualizing interprofessional teams as multi-team systems—Implications for assessment and training. *Teach Learn Med*. 2015;27(4):366–9.
142. West M, Armit K, Loewenthal L, Eckert R, West T, Lee A. Leadership and leadership development in healthcare: The evidence base. London: Faculty of Medical Leadership and Management; 2015.
143. West MA, Lyubovnikova J. Illusions of team working in health care. *J Health Organ Manag*. 2013;27(1):134–42.
144. Wierenga D, Engbers LH, Van Empelen P, De Moes K, Wittink H, Gründemann R, et al. The implementation of multiple lifestyle interventions in two organizations: a process evaluation. *J Occup Environ Med*. 2014;56(11):1195.
145. Wierenga D, Engbers LH, Van Empelen P, Duijts S, Hildebrandt VH, Van Mechelen W. What is actually measured in process evaluations for worksite health promotion programs: a systematic review. *BMC Public Health*. 2013;13(1):1.
146. Wierenga D, Engbers LH, Van Empelen P, Hildebrandt VH, Van Mechelen W. The design of a real-time formative evaluation of the implementation process of lifestyle interventions at two worksites using a 7-step strategy (BRAVO@ Work). *BMC Public Health*. 2012;12(1):619.
147. Wierenga D, Engbers LH, Van Empelen P, van Mechelen W. A 7-Step Strategy for the Implementation of Worksite Lifestyle Interventions: Helpful or Not. *J Occup Environ Med*. 2016;58(5).

148. Wildman JL, Thayer AL, Rosen MA, Salas E, Mathieu JE, Rayne SR. Task types and team-level attributes: Synthesis of team classification literature. *Human Res Develop Rev.* 2012;11(1):97–129.
149. Wilkinson J, Powell A, Davies H. Are clinicians engaged in quality improvement? A review of the literature on healthcare professionals' views on quality improvement initiatives 2011. <http://www.health.org.uk/sites/health/files/AreCliniciansEngagedInQualityImprovement.pdf>. Accessed 26 Mar 2020.
150. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. 18th international conference on evaluation and assessment in software engineering; London2014.
151. Wong AH-W, Gang M, Szyld D, Mahoney H. Making an “Attitude Adjustment”: Using a Simulation-Enhanced Interprofessional Education Strategy to Improve Attitudes Toward Teamwork and Communication. *Simul Healthc.* 2016;11(2):117–25.
152. Zhu X, Baloh J, Ward MM, Stewart GL. Deliberation makes a difference: Preparation strategies for TeamSTEPPS implementation in small and rural hospitals. *Med Care Res Rev.* 2015;73(3):283–307.
153. Zwarenstein M, Goldman J, Reeves S. Interprofessional collaboration: effects of practice-based interventions on professional practice and healthcare outcomes. *Cochrane Database Syst Rev.* 2009(3); doi:10.1002/14651858.CD000072.pub2.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors declare that they have obtained no funding for this study.

Authors' contributions

WK conceived and designed the study, WK and RW performed data acquisition and analysis, WK drafted the work, with substantial contributions of RW and CW. CW latter also substantively revised the final version. The submitted version of this paper was approved by ALL authors.

Acknowledgements

Not applicable.

Figures

Figure 1

Yearly total trends of indexed (PubMed) publications on 'team training' or 'crew resources management' (period: 1990 - 2019)

Figure 2

Phased development of the ReCoMuTe checklist (some arrows depict feedback loops)

Figure 3

Flowchart of the literature-selection process