

# Do team processes really have an effect on clinical performance? A systematic literature review

J. Schmutz\* and T. Manser

Industrial Psychology and Human Factors, Department of Psychology, University of Fribourg, Rue de Faucigny 2, 1700 Fribourg, Switzerland

\* Corresponding author. E-mail: jan.schmutz@unifr.ch

## Editor's key points

- This review has examined the impact of team process behaviours on clinical performance.
- Twenty-eight studies, which reported at least one relationship between team process or an intervention and outcome, were reviewed.
- Team process behaviours have been shown to influence performance.
- Training in team behaviours results in improved performance.

**Summary.** There is a growing literature on the relationship between team processes and clinical performance. The purpose of this review is to summarize these articles and examine the impact of team process behaviours on clinical performance. We conducted a literature search in five major databases. Inclusion criteria were: English peer-reviewed papers published between January 2001 and May 2012, which showed or tried to show (i) a statistical relationship of a team process variable and clinical performance or (ii) an improvement of a performance variable through a team process intervention. Study quality was assessed using predefined quality indicators. For every study, we calculated the relevant effect sizes. We included 28 studies in the review, seven of which were intervention studies. Every study reported at least one significant relationship between team processes or an intervention and performance. Also, some non-significant effects were reported. Most of the reported effect sizes were large or medium. The study quality ranged from medium to high. The studies are highly diverse regarding the specific team process behaviours investigated and also regarding the methods used. However, they suggest that team process behaviours do influence clinical performance and that training results in increased performance. Future research should rely on existing theoretical frameworks, valid, and reliable methods to assess processes such as teamwork or coordination and focus on the development of adequate tools to assess process performance, linking them with outcomes in the clinical setting.

**Keywords:** clinical competence; group processes; leadership; patient care team; patient safety

Breakdown in team processes such as coordination, leadership, or communication have frequently been associated with adverse events and patient harm<sup>1–3</sup> and the effectiveness of such team processes is central to the successful provision of patient care.<sup>1 4 5</sup> While recent reviews indicate that team processes are widely accepted as an important factor influencing clinical performance of medical teams,<sup>1 5–8</sup> a general framework is needed to classify and compare different studies on teamwork. In this review, we invoked McGrath's systemic input–process–output (IPO) framework<sup>9</sup> that has served as a foundation for numerous studies in team research<sup>10–14</sup> and has been adapted and used in clinical settings in recent years.<sup>5 7 15–17</sup>

According to this framework, *inputs* are preconditions influencing the processes in the team (e.g. team climate, task structure, leadership style). *Team processes* are defined as the cognitive, verbal, and behavioural activities going on while the team is working together (i.e. team communication, team leadership, team coordination, and team decision-making).<sup>5 18 19</sup> *Outputs* are the product of these processes. Either patient outcomes or team outcomes can be considered as *outputs* in a clinical setting.<sup>5</sup>

The IPO framework conceptualizes performance as an output that is directly influenced by team processes,<sup>5 9</sup> but does not provide explicit definitions of performance or a means by which to measure it. Various authors agree that there is both a process and an outcome-related aspect to performance.<sup>20–22</sup> The distinction between outcome and process performance measures is not always consistently used in the literature but should be borne in mind when aiming to establish an empirical evidence base on the relationships between team processes and outcomes.

Outcome performance measures such as mortality,<sup>23</sup> morbidity,<sup>23</sup> or length of stay<sup>24</sup> can be assessed objectively without consideration of the team process. Process performance measures, in contrast, are action-related aspects of performance embedded in the team processes.<sup>15</sup> Process performance measures are often more easily accessible and less influenced by other variables than outcome performance measures because they refer to directly observable behaviours executed by the team during patient treatment (e.g. measuring task execution time, rating specific behaviours according to medical guidelines).<sup>25 26</sup>

In the infancy of team research in medicine, the main aim was to generate a general understanding of which team processes influenced performance in which way. After qualitative studies investigating which team processes might be relevant to clinical performance,<sup>27 28</sup> quantitative studies were conducted to develop a clearer understanding of the impact of team processes on clinical performance. Studies investigated the association between team processes and either process performance<sup>7 29</sup> or outcome performance measures.<sup>23</sup> However, despite this improved understanding, it is still not clear how large the effect of these relationships is because in the majority of cases, no effect sizes are reported.

This systematic literature review aims to address this gap by analysing articles that investigate the relationship between team processes and clinical performance measures (i.e. process or outcome performance) and to report and compare the respective effect sizes. Furthermore, we will describe and discuss the different team processes and clinical performance measures used. This knowledge is needed to design targeted studies and effective interventions for patient care teams.

## Methods

We conducted a literature search based on the recommendations of the PRISMA statement<sup>30</sup> consulting the databases PubMed, Science Direct, PsycINFO, PSYINDEXplus Literature, and Audiovisual Media. Additionally, a meta-search with Google Scholar was conducted; of which, only the first 50 results were examined. The search term used was PATIENT SAFETY combined with TEAMWORK, COMMUNICATION, or LEADERSHIP. In addition, a hand search was conducted based on the references of the identified articles. The literature search was conducted in May 2012.

Figure 1 provides an overview of the inclusion criteria and the five-step selection procedure. We selected English articles published in journals between January 2001 and May 2012 investigating the relationship between team processes and clinical performance. We selected articles that showed or tried to demonstrate (i) a statistical relationship between a team process variable and clinical performance (process or outcome performance) or (ii) an improvement of clinical performance (process or outcome performance) through an intervention targeting team processes.

We included only articles with performance measures. We excluded articles which used self-report data because surveys or interviews about the teams' own perception of performance can contain a self-report bias<sup>31</sup> and could potentially have distorted the results of this review. Intervention studies were only considered when targeting a team process behaviour (e.g. through training) and not implying structural changes (e.g. care pathways)<sup>32</sup> at the same time, because this would preclude distinguishing between effects of the training vs the structural change. We included studies using process or outcome performance measures. Since our main focus was on factors influencing patient

care, we excluded studies measuring team outcomes (e.g. job satisfaction, stress, burnout).<sup>5</sup>

Each step was performed independently by two reviewers (J.S. and Mariel Dardel). The agreement was between 90% and 94% in each step. Any disagreement in the selection process was resolved by extensive discussion.

## Rating of study quality

In order to assess the methodological quality of the selected articles, we used a rating system based mainly on the one proposed by Buckley and colleagues.<sup>33</sup> Since external validity is an important quality indicator, we replaced the single item by Buckley and colleagues with two items from a checklist by Downs and Black.<sup>34</sup> For intervention studies, three items concerning the quality of the intervention were added from Downs and Black. The question of triangulation was not applied to the intervention studies because the focus was on the effect of the intervention and we did not expect authors of intervention studies to triangulate multiple methods. The complete list and a detailed description of quality indicators can be found in Supplementary Table S3.

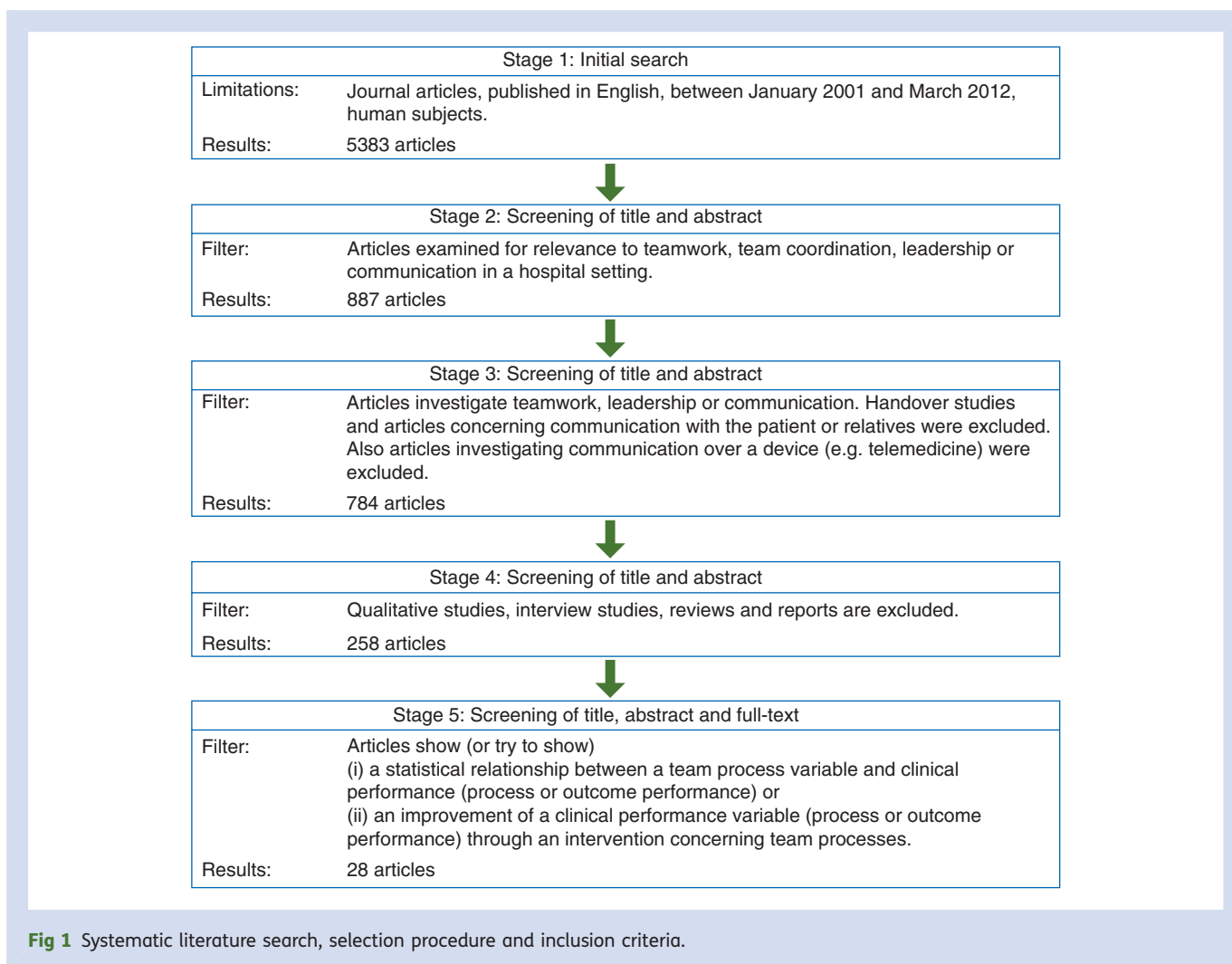
Each indicator was scored as '0' (not fulfilled), '0.5' (partially fulfilled), '1' (complete), or 'not mentioned' (i.e. information not explicitly provided and thus unclear whether the criterion has been fulfilled or not). Quality ratings were performed by J.S. A random sample of five studies was rated by T.M. We achieved consistency of 91%. Disagreements in the ratings were due to different interpretations of the descriptions in the articles and were resolved by discussion.

## Data extraction

The following characteristics of the selected studies that were deemed most relevant were extracted, to evaluate the statistical relationships between team processes and clinical performance: team process behaviours, performance measures, participants, and results plus a description of the intervention in the case of intervention studies. Additionally, we calculated the effect size for every statistical process–output relationship reported in the selected studies based on the data provided in the articles. This enabled us to determine not only if team processes are significantly related with clinical performance but also how large this effect is and if it is large enough to be relevant for practical implications.<sup>35</sup> We report only significant and non-significant effects that were explicitly stated in the selected articles, although additional relationships may have been investigated but not reported.

## Results

As can be seen from Figure 1, the initial search yielded 5383 articles. After excluding the irrelevant studies in stage 2, 887 articles remained. In stage 3, 784 studies were selected, of which 258 used quantitative methods and were retained for stage 4. After applying the final selection step, we identified 28 studies; of which, seven were intervention studies. Table 1 and Table 2 provide an overview of the relevant



characteristics pertaining to all the articles included in this review.

**Team processes investigated and their measurement**

The selected studies examined various team processes: communication,<sup>23 24 36-38</sup> coordination,<sup>24 39-41</sup> leadership,<sup>7 24 31 42 43</sup> non-technical skills,<sup>29 44-49</sup> team behaviour,<sup>42</sup> team monitoring behaviour,<sup>50</sup> and teamwork.<sup>23 36 51</sup> Six studies examined more than one team process behaviour.<sup>23 24 31 36 42 43</sup>

In reviewing the articles, we noted a high variability in the research approaches and measures used to study these team processes. As can be seen from Table 3, observational studies were most prominent. Most studies used video-based behaviour coding of data obtained in a simulator setting (n=10). Of the nine studies conducted in a clinical setting, three used video-based and six used live behaviour coding. Only three studies used surveys to collect team process data.

At the measurement instruments level, we found that four of the seven studies examining non-technical skills used the Surgical NOTECBS system.<sup>29 44-46</sup> The other three systems used were the Behavioural Marker Risk Index (BMRI),<sup>47</sup> the Anaesthetists' Non-Technical Skills (ANTS),<sup>49</sup> and one specific

behavioural marker system for neonatal resuscitation.<sup>48</sup> Three of the six studies investigating communication used different observation systems<sup>31 36 38</sup> and the other three all used different questionnaires.<sup>23 24 37</sup> Three studies conceptualized the team processes under investigation as teamwork. Of these studies, one used the Safety Attitude Questionnaire (SAQ),<sup>23</sup> one used a rating system for teamwork behaviour,<sup>51</sup> and one study focused on events disrupting teamwork.<sup>36</sup> Of the five studies investigating leadership processes, four conducted observations but used different observation systems<sup>31 42 43 52</sup> and one study used a survey.<sup>24</sup> Of the four studies focusing on coordination, three<sup>39-41</sup> used the coding system of Manser and colleagues<sup>53</sup> and one assessed coordination using a survey.<sup>24</sup>

**Process and outcome measures of clinical performance**

Table 4 summarizes the 50 performance measures used in the 28 studies sorted into 41 process performance measures and nine outcome performance measures. Fourteen studies recorded deviations (i.e. errors, problems, or non-routine events during treatment) as a measure of process

**Table 1** Characteristics of studies reporting relationships between team process behaviour and process or outcome performance

Study	Team process behaviour / research method and tool	Performance measure / method	Participants / setting	Results	Effect size*	Quality score (max = 12)	'Not mentioned's
Burtscher and colleagues <sup>39</sup>	<i>Team coordination / behaviour coding of video data</i> → coding system for coordination <sup>53</sup> consisting of 33 codes, which are grouped into five main categories: information management, task management, coordination via work environment, metacoordination, and other communication	Clinical performance of the anaesthesia induction / checklist-based rating system by experts	Anaesthesia staff, 19 anaesthetists, 14 nurses, teams of 2-4 persons / clinical setting (22 videos of routine anaesthesia inductions)	High performing teams show a more pronounced increase in task management in response to NRE in contrast to low-performing teams	Low performing teams: $\bar{x}_1 \approx 23\% \dagger$ (routine) to $\bar{x}_2 \approx 29\% \ddagger$ (NRE) vs. high performing teams $\bar{x}_3 \approx 16\% \dagger$ (routine) to $\bar{x}_4 \approx 36\% \ddagger$ (NRE) (relative amount of time teams spent on task management); $t(20) = -2.75, p < .05$	10	1
Burtscher and colleagues <sup>40</sup>	<i>Adaptive coordination while different phases of a treatment / behaviour coding of video data</i> → coding system for coordination <sup>53</sup> consisting of 33 codes, which are grouped into five main categories: information management, task management, coordination via work environment, metacoordination, and other communication	Decision latency / time from the recognition of the asystole until the decision how to respond to it Execution latency / time from deciding what to do until restoration of sinus rhythm)	15 anaesthesia trainees, 15 anaesthesia nurses, teams of 2 persons / simulation (standard anaesthesia induction)	Negative association between decision latency and the anaesthesia trainees change in information management No association between other coordination aspects and decision latency or execution time	$r = -.49$ ( $p = .003$ ) NS	9	1
Burtscher and colleagues <sup>50</sup>	<i>Team monitoring behaviour / behaviour coding of video data</i> → coding each time a team member was observing the action of a teammate	Clinical performance of the anaesthesia induction / checklist-based rating system by experts	31 anaesthesia resident, 31 anaesthesia nurses, teams of 2 persons / simulation (anaesthesia induction)	Negative association between team monitoring and performance	$r = -.44$ ( $p = .02$ )	10	1
Carlson and colleagues <sup>42</sup>	<i>Leadership (LS) and team behaviour / behaviour coding of video data</i> → global assessment of one dominant style of leadership (transactional LS, flexible/dynamic team LS, neither); rating (0-4) of four team behaviour categories (workload management, communication, prioritizing and reassessing priorities, vigilance)	Standard of care / expert assessment in consideration of behavioural guidelines (poor, marginal, standard of care)	113 3 <sup>rd</sup> - year undergraduate medical students, teams of 2-3 persons / simulation (acute dyspnea)	Pos. association of the average team score (mean of the four dimensions) and standard of care No interrelation of LS style and standard of care	$r = .77$ ( $p < .0001$ ) NS	8	2

Catchpole and colleagues <sup>44</sup>	<i>Non-technical skills / behaviour coding of video and live data</i> → Surgical NOTECHS measurement framework	Problems / observation Intraoperative performance / checklist-based rating system by experts Operating time	42 paediatric and orthopaedic operation teams / clinical setting (paediatric and orthopaedic operations)	Teams with effective teamwork have: Fewer minor problems per operation Higher intraoperative performance Shorter operating times than teams with less effective teamwork	† $\bar{x}_1 \approx 7.3$ † (effective teams) vs. $\bar{x}_2 \approx 11.1$ † (ineffective teams) $t=3.05$ $p=.004$ † $\bar{x}_3 \approx 93.3\%$ † (effective teams) vs. $\bar{x}_4 \approx 95.5\%$ † (ineffective teams) $t=-3.25$ $p=.002$ † $\bar{x}_5 \approx 195\text{min.}$ † (effective teams) vs. $\bar{x}_6 \approx 153\text{min.}$ † (ineffective teams) $t=2.25$ $p=.03$	9.5	1
Catchpole and colleagues <sup>73</sup>	<i>Non-technical skills / live behaviour coding</i> → Surgical NOTECHS measurement framework	Operating time Errors in surgical technique / observation Other procedural problems and errors / observation	48 surgical operation teams / clinical setting (26 laparoscopic cholecystectomies, 22 carotid endarterectomies)	Association between situation awareness and errors in surgical technique Association between LS & Management and operating time Association between LS & Management score of the nurse and other procedural problems and errors No association between other NOTECHS dimensions and any performance measure	† (F(2,42)=7.93, $p=0.001$ ) † (F(2,42)=3.32, $p = 0.046$ ) † (F(5,1)=3.96, $p=0.027$ ) NS	9.5	1
Davenport and colleagues <sup>23</sup>	<i>Teamwork and communication / survey</i> → Safety attitudes questionnaire (SAQ)	Mortality (patient death in or out of the hospital 30 days after the operation) / data from the National Surgical Quality Improvement Program (NSQIP) Morbidity (patient having 1 or more postoperative complications up to 30 days after operation) / Data from the NSQIP)	6083 staff members of general and vascular surgery from 44 Veterans Affairs and 8 academic medical centres / clinical setting	Significant negative correlation between morbidity and Positive communication of surgical service care providers with attending doctors Positive communication of surgical service care providers with residents No interrelation of teamwork and mortality No interrelation of teamwork and morbidity	$r = -.38$ ( $p < 0.01$ ) $r = -.25$ ( $p < 0.08$ ) NS NS	9	1

Continued

Table 1 Continued

Study	Team process behaviour / research method and tool	Performance measure / method	Participants / setting	Results	Effect size*	Quality score (max=12)	'Not mentioned's
ElBardissi and colleagues <sup>36</sup>	<i>Teamwork and communication disruptions</i> / live behaviour coding → any occurrence concerning teamwork /communication that disrupted the flow of the operation	Errors (events failed its intended outcome) / observation	5 surgeons / clinical setting (31 cardiac surgical operations)	Positive association between teamwork disruptions and surgical errors	$r = .67$ ( $p < 0.001$ )	11	0
Künzle and colleagues <sup>7</sup>	<i>Leadership</i> / behaviour coding of video data (structuring LS, content oriented LS and total amount of LS)	Execution time	12 anaesthesia teams / simulation (anaesthesia induction)	Negative association between execution time during routine and highly standardized phases and structuring LS and content oriented LS and total amount of LS No significant association between LS and execution time during a nonroutine event	$r = -.59$ ( $p < .05$ ) $r = -.52$ ( $p < .10$ ) $r = -.56$ ( $p < .05$ ) NS	9.5	1
Manojlovich and colleagues <sup>37</sup>	<i>Communication</i> / survey → ICU Nurse-Physician Questionnaire (4 scales: openness, accuracy, timeliness and understanding)	Ventilator-associated pneumonia / data from the hospital database Bloodstream infections / data from the hospital database Pressure ulcers / data from the hospital database	462 nurses from 25 ICUs / clinical setting	Negative association between timeliness and pressure ulcers. No significant association between overall communication or other subscales and outcome variables	$r = -.38$ ( $p = .06$ ) NS	9.5	1
Manser and colleagues <sup>41</sup>	<i>Team coordination</i> / behaviour coding of video data → coding system for coordination <sup>53</sup> consisting of 33 codes, which are grouped into five main categories: information management, task management, coordination via work environment, metacoordination, and other communication	Clinical performance / checklist-based rating system by experts	48 first year students, teams of 2 persons / simulation (malignant hyperthermia)	Positive association between task distribution and performance	$r = -.466$ ( $p < 0.01$ )	10	1

Marsch and colleagues <sup>43</sup>	<i>Leadership, task distribution and information transfer / behaviour coding of video data if the specific behaviour is present or not</i>	Clinical performance / time-based scoring system for critical treatment steps	16 teams consisting of 2 nurses and 1 physician each / simulation (cardiopulmonary resuscitation)	Successful teams show more task distribution and  more LS behaviour than failing teams.  No significant difference in information transfer	6/6 (successful teams showed task distribution) vs. 4/10 (failing teams showed task distribution), odds ratio can't be calculated  OR=8 (successful teams show 8 times more likely LS behaviour than failing teams  NS	8.5	2
Mazzocco and colleagues <sup>47</sup>	<i>Non-technical skills / live behaviour coding according to the behaviour marker risk index (BMRI)</i>	Outcome score (1= no complications to 5=death or permanent disability) including complications and other significant postoperative outcomes / retrospective chart review of the concerning patients	130 physicians, nurses, operating room technicians, nurse anaesthetists / clinical setting (300 surgical cases)	Patients were more likely to experience death or a major complication when there were less teamwork behaviours	OR=4.82 (corrected for preoperative physical fitness)	10	1
Mishra and colleagues <sup>29</sup>	<i>Non-technical skills / live behaviour coding → Surgical NOTECHS measurement framework</i>	Technical errors / observation	Surgeons, anaesthetists, nurses / clinical setting (26 elective laparoscopic cholecystectomy operations)	Negative association between situation awareness of surgeons and technical errors	$\rho = -.718$ ( $p < .001$ )	10	1
Pollack and colleagues <sup>24</sup>	<i>Leadership, communication and coordination / survey → organizational assessment tool</i>	Mortality Bronchopulmonary dysplasia Periventricular / intraventricular haemorrhage or leukomalacia Retinopathy of prematurity Length of stay  All outcomes were collected from clinical records	Nurses, physicians and respiratory therapists of 8 neonatal intensive care units / clinical setting (493 deliveries)	Positive association between leadership and PIVH/PVL Negative association between coordination and PIVH/PVL  No significant results for communication and the other outcome measures	$\dagger p < .001$ $\dagger p = .047$  NS	10.5	0
Schraagen <sup>74</sup>	<i>Non-technical skills / live behaviour coding → Surgical NOTECHS system</i>	Non-routine events/ observation Operating time 30-day postsurgical outcome (uncomplicated, minor complication, major complications or death)	Paediatric cardiac surgical teams / clinical setting (40 operations)	Positive association between teamwork and operating times  No association between teamwork and non routine event  No association between teamwork and outcome	$r = .45$ ( $p < 0.01$ )  NS  NS	10	1

Continued

Table 1 Continued

Study	Team process behaviour / research method and tool	Performance measure / method	Participants / setting	Results	Effect size*	Quality score (max=12)	'Not mentioned's
Siassakos and colleagues <sup>51</sup>	Teamwork / behaviour coding of video data → Generic teamwork score (GTS)	Clinical efficiency score / check-list rating Time until turning the patient into the recovery position Time until administration of O <sub>2</sub> Time until venous blood sampling	24 teams consisting of 2 doctors and 4 midwives each / simulation (obstetric emergency)	Positive association between clinical efficiency score and GTS Negative association between GTS and time until turning the patient into the recovery position time until administration of O <sub>2</sub> time until venous blood sampling	$r=.72^{\#}$ ( $p<0.001$ )  $r=-.38^{\#}$ ( $p=0.026$ )  $r=-.52^{\#}$ ( $p=0.002$ )  $r=-.60^{\#}$ ( $p<0.001$ )	9.5	1
Thomas and colleagues <sup>48</sup>	Non-technical skills / behaviour coding of video data → behaviour marker system	Compliance with Neonatal Resuscitation Program (NRP) guidelines / checklist-based rating system by experts	Neonatal resuscitation teams consisting of two providers, one physician one neonatal nurse / clinical setting (132 deliveries)	Negative correlation between total NRP noncompliance and Communication Management No correlation between leadership and total NRP noncompliance	$\rho=-.021$ , $p=0.014$ $\rho=-.020$ , $p=0.021$ NS	10	1
Tschan and colleagues <sup>31</sup>	Directive leadership and structuring inquiry / behaviour coding of video data → coding system derived from the guidelines for cardiopulmonary resuscitation and other observational systems	Medical performance / checklist based rating of technical acts Percentage of time the patient did not show signs of normal circulation and received cardiovascular support	21 teams consisting of 3 nurses, 1 resident, 1 senior doctor / simulation (cardiac arrest)	Positive association between performance and directive LS of the first nurse and structuring inquiry of the nurses and directive LS of the resident (in the first 30s when he enters the room) and structuring inquiry of the senior physician	$r=.445$ ( $p<.05$ )  $r=.216$ ( $p<.05$ )  $r=.522$ ( $p<.01$ )  $r=.428$ ( $p<.01$ )	10	1
Tschan and colleagues <sup>38</sup>	Explicit reasoning and talking to the room / behaviour coding of video data → sum of reasoning units, talking to the room present or not (dummy variable)	Diagnostic accuracy / evaluation of the team diagnosis	20 Groups consisting of 2 or 3 experienced physicians / simulation (anaphylactic shock)	Successful teams show more explicit reasoning (# of linked utterances) and more talking to the room than less successful teams No significant difference in the amount of information	$\dagger 4.0$ (successful) $> 1.13$ (successful with help) $> 1.0$ (fail) $F(2,15)=5.750$ ; $p=.014$ $\hat{w}^2=0.43$  NS	10.5	1



Westli and colleagues <sup>49</sup>	Non-technical skills / behaviour coding of video data → ANTS system (revised)	-Performance score / checklist-based rating system by experts -Medical Management / overall rating from 1-5	27 trauma teams consisting of 1 surgeon 1 anaesthesiologist, 2 nurses, 1 radiographer / simulation (resuscitation)	considered of successful and less successful teams		9.5	1
				Negative association between Medical Management and poor coordination	$r = -.36$ ( $p < 0.01$ )		
				Performance score and supporting behaviour	$r = -.37$ ( $p < 0.01$ )		
				Positive association between Medical Management and information exchange	$r = .34$ ( $p < 0.01$ )		
				No correlation between performance score and coordination	NS		
				poor Coordination	NS		
				information exchange	NS		
				use of authority	NS		
				poor use of authority	NS		
				assessing capabilities	NS		
				poor supporting behaviour	NS		
				No correlation between Medical Management and coordination	NS		
				use of authority	NS		
				poor use of authority	NS		
				assessing capabilities	NS		
				supporting behaviour	NS		
				poor supporting behaviour	NS		

\*  $r$ ,  $\rho$  and  $\hat{w}^2$  effect sizes are interpreted as follows:  $r$  or  $\rho = .10$  small effect;  $r$  or  $\rho = .30$  . medium effect;  $r$  or  $\rho = .50$  large effect<sup>62, 75</sup>;  $\hat{w}^2 = 0.01$  small effect,  $\hat{w}^2 = 0.09$  medium effect,  $\hat{w}^2 = 0.25$  large effect<sup>73</sup>

† The required information to calculate the effect sizes are not available. If available the absolute sizes are indicated instead.

‡ Means are assessed out of figures. The exact means are not mentioned in the text.

¶ Kendall's Tau ( $\tau$ ) was transformed into  $r$  according to Walker<sup>74</sup>

§ 'Not mentioned' means it was unclear if something has been done or not based on the information provided in the article. NS, Not Significance.

**Table 2** Characteristics of team process behaviour interventions and their impact on performance. Team process measures used in the intervention studies are not listed here because for these studies, the focus is on the effect of the intervention on performance and not on the process. \*The required information to calculate the effect sizes was not available. The absolute sizes are indicated instead. <sup>†</sup> $\hat{w}^2$  effect sizes are interpreted as follows:  $\hat{w}^2 \geq 0.01$  small,  $\hat{w}^2 \geq 0.09$  medium, and  $\hat{w}^2 \geq 0.25$  large;<sup>71</sup> Cohen's *d* effect sizes are interpreted as follows:  $d \geq 0.20$  small,  $d \geq 0.50$  medium, and  $d \geq 0.80$  large.<sup>71</sup> †'Not mentioned' means it was unclear if something has been done or not based on the information provided in the article

Study	Intervention/design/team process measure*	Performance measure/method	Participants	Results	Effect size <sup>†</sup>	Quality score (max=14)	'Not mentioned' <sup>†</sup>
Fernandez Castelao and colleagues <sup>50</sup>	Video based crew resource management training/quasi-experimental control group post-test design/no team process measure	No-flow time (time with no chest compression)	Four-person medical student teams, 26 teams in the experimental group, 18 teams in the control group	Less no flow time in the post-intervention group comparing with the control group	$\bar{x}_1 = 36.3\%$ (control) vs $\bar{x}_2 = 31.4$ (experimental) ( $P=0.014$ )*	11	0
Kalisch and colleagues <sup>54</sup>	Staff teamwork and engagement enhancement intervention/quasi-experimental uncontrolled pre-test-post-test design/post-interview about teamwork	Fall rates per 1000 patient days/information from patient report	49 nurses, six unit secretaries of a community hospital	Patient fall rates decreased after the intervention	$\bar{x}_1 = 7.73$ to $\bar{x}_2 = 2.99$ falls per 1000 patient days ( $t=3.98$ , $P<0.001$ )*	9	2
McCulloch and colleagues <sup>55</sup>	Intervention based on principles of civil aviation crew resource management/quasi-experimental uncontrolled pre-test-post-test design/NOTECHS and SAQ teamwork climate score	Operating technical errors/ observation Operating time Length of stay	Surgeons, anaesthetists, nurses performing 48 operations in the pre-intervention group and 55 operations in the post-intervention group	Less operating technical errors after the intervention No reduction in operating time No reduction in length of stay	$d=0.63$ ( $P=0.009$ ) NS NS	10.5	0
Morey and colleagues <sup>57</sup>	Emergency Team Coordination Course (ETCC)/quasi-experimental control group design with one pre- and two post-tests/Behaviour Anchored Rating Scales (BARS)	Errors/observation	Physicians, nurses, and technicians of six emergency departments (EDs) in the experimental group ( $n=684$ ) and three EDs in the control group ( $n=374$ )	Decrease in the clinical error rate in the post-intervention group No significant difference between the experimental and control group	$d=1.93$ ( $P=0.039$ ) NS	12	1
Nielsen and colleagues <sup>58</sup>	MedTeams Labor and Delivery Team Coordination Course based on crew resource management trainings/cluster-randomized control group design with no pre-test/no team process measure	Adverse maternal Outcome Index (number of patients with one or more adverse outcomes divided by the total number of deliveries)/information from patient report	Obstetrician, anaesthesiologist, and nurses of seven hospitals (obstetrics) in the experimental group ( $n=1307$ ) and eight hospitals in the control group	Significant reduction in Caesarean delivery decision to incision	$\bar{x}_1 = 33.3$ min (control) vs $\bar{x}_2 = 21.2$ (experimental) ( $P=0.039$ )*	12	0

<p>Immediate Caesarean delivery decision to incision (min)/observation</p>	<p>NS</p> <p>No reduction in adverse maternal outcomes in the experimental group</p>	<p>1</p>
<p>Adaptation of a crew resource management training with a simulator/quasi-experimental uncontrolled pre-test – post-test design/safety culture survey from the Agency for Healthcare Research and Quality (AHRQ) consisting a teamwork and communication scale</p>	<p>Adverse Outcomes Index (AOI)/report</p> <p>Nurse midwives, gynaecologist physicians, residents, anaesthetists, and unit secretaries of a labour and delivery unit with ~9200 births per year</p>	<p>9.5</p> <p>AOI index decreases significantly from <math>\bar{x}_1 = 0.052</math> to <math>\bar{x}_2 = 0.043</math> (no P-value stated)*</p>
<p>Medical team training (CRM based)/quasi-experimental uncontrolled design with one pre- and two post-tests/SAQ (teamwork climate scale)</p>	<p>Operating teams consisting of surgeons, residents, anaesthesiologists, and nurses of a hospital with eight operating theatres</p>	<p>11</p> <p>Cases with any delay significantly decreased from <math>\bar{x}_1 = 23.2\%</math> to <math>\bar{x}_2 = 10\%</math> (<math>P &lt; 0.0001</math>)*</p>

**Table 3** Methods and research settings for studying the team process – performance relationship

Methods/settings	Number of articles	Study reference number
Live observation/clinical setting	6	29,36,44,47,45,72
Video-based observation/clinical setting	3	39,44,48
Video-based observation/simulation	10	7,31,38,40–43,49–51
Survey/clinical setting	3	23,24,37

performance. Performance checklists based on clinical guidelines were the next frequently used performance measure ( $n=10$ ) followed by the time until a specific treatment is conducted ( $n=7$ ). The outcome performance measure used most frequently was complications after treatment ( $n=4$ ).

### Effects of team processes on performance in the non-intervention studies

In total, the 21 studies reported 66 relationships of a team process variable with a performance variable. Forty of these effects were significant and 26 were non-significant. Thirteen of the 21 non-intervention studies calculated correlations to investigate this relationship. More than one performance measure was used by 15 studies and 12 of these reported both non-significant and significant effects. Only six studies investigated just one effect and assessed only one performance measure. All of them were significant.

No study explicitly reported effect sizes. The effect sizes calculated are shown in Table 1 and Table 2. They range from very high ( $r=0.77$ )<sup>42</sup> to small ( $\rho=-0.02$ ).<sup>48</sup> Only one study reported a small effect,<sup>48</sup> while all the others described effects considered as large or medium.

### Interventions targeting team process behaviours

The interventions were carried out in community hospitals,<sup>54</sup> operating theatres,<sup>55 56</sup> emergency departments,<sup>57</sup> and labour and delivery units.<sup>58 59</sup> Five of the seven intervention studies used training explicitly based on crew resource management (CRM) principles,<sup>55 56 58–60</sup> while the other two studies included some CRM elements such as an introduction to teamwork and non-technical skills. According to the brief descriptions in the articles, it appears that all interventions were of similar content. Typical topics discussed in the training were principles of teamwork and human factors, situation awareness, improvement of team skills, communication, and leadership. The duration of the training ranged from 1 to 2 days and included methods such as theoretical lectures on CRM principles, video analysis, and role-playing. Unfortunately, an exact comparison of the interventions is not possible due to the limited descriptions of the training provided in the articles. Table 2 summarizes the effects of the seven interventions, all

**Table 4** Performance measures used. \*If multiple performance measures are used in one article, the study is mentioned several times. NREs, non-routine events

Performances measures used	Total number of performance measures used	Study reference number*
Process performance measure		
Deviations (errors, problems, NREs during the treatment)	14	24,24,24,29,36,37,37,37,44,44,55,57,45,72
Case delays	1	56
Length of stay	2	24,55
Operating time	5	7,44,55,45,72
Percentage of time the patient receives a specific treatment	2	31,60
Time until a specific treatment is conducted	7	40,40,43,51,51,51,58
Performance checklists	10	31,39,41,42,44,48,49,49,50,51
Outcome performance measures		
Complications after operation	4	47,58,59,72
Diagnostic accuracy	1	38
Fall rates	1	54
Morbidity	1	23
Mortality	2	23,24

indicating significant improvements of performance after the intervention. The intervention studies reported 11 effects on a performance measure; of which, seven were significant. Three studies assessed more than one performance measure. Only two studies indicated all the information to calculate the effect size and they reported one medium<sup>55</sup> and one large effect.<sup>57</sup>

### Quality of the selected studies

A complete list of the quality ratings for every article can be found in the Supplementary Table S1 and Supplementary Table S2. The study quality ratings ranged from 9 to 12 points out of 14 for the intervention studies and from 8 to 11 out of 12 points for the other studies. Overall, data collection methods were found to be reliable and valid to answer the specific research questions. Two common problems were the poor discussion of potential confounding factors and the use of a single data collection method instead of strengthening the results through triangulation.

All non-intervention studies were prospective. In general, research questions were clearly stated, methods well described, analyses were appropriate, and the conclusions clearly justified by the results.

All intervention studies used quasi-experimental or clustered designs. Only three of the seven intervention studies applied a control group design, while the other four were pre-test-post-test studies. Two studies included a follow-up post-test to investigate long-term effects. All intervention studies provided unspecific descriptions of the conducted interventions limiting their reproducibility.

### Other study characteristics

The studies included participants of various professions examining teams consisting of anaesthetists, nurses, medical

students, paediatricians, surgeons, operating theatre technicians, and midwives. In four studies, the participants were uni-professional.<sup>37 38 41 42</sup>

## Discussion

The aim of our systematic literature review was to consolidate the statistical evidence for the effects of team processes on clinical performance in patient care teams. Furthermore, we provide an overview of all team process and performance measures used in these studies that will inform future research in this field regarding the strength and weaknesses of current measures and necessary developments.

Focusing on the process-performance relationship, this review found that significant progress has been made in recent years. Most studies report strong effects indicating that team processes are significantly influencing clinical performance. However, we identified areas for improvement with regard to defining and measuring both team processes and clinical performance. Our systematic analysis of study quality also points at possible improvements in both study design and reporting.

Most studies did not refer to a conceptual framework. They sometimes used vague definitions of the two concepts 'team process behaviours' and 'performance' and a broad range of measurement approaches was also seen. An appropriate scientific definition and explicit reference to a common conceptual framework are prerequisites for comparing studies that investigate a broad spectrum of team process behaviours. Such a framework aids in study design and interpretation of results. Although the IPO model is rather simple, it is widely accepted and has proven useful in various teamwork settings. The IPO model facilitates the research process by providing a clear structure of potential

relationships upon which to focus (e.g. the impact of team mental models as an input on team process behaviours such as decision-making or the relationship between leadership processes within the team and subjective outcomes such as staff well-being).

While more complex models such as the input-throughput-output model of team adaptation of Burke and colleagues<sup>61</sup> have been developed to reflect the complexity of teamwork, these models are often too complex for isolating research questions that can be tested in an actual work setting. We have to strive for a balance between complexity and feasibility for these models to be useful in guiding team research in healthcare and in conceptually clarifying the relevant inputs, team process behaviours, and outcomes.

Most studies measuring team process behaviours have used observational methods. This is a more time-consuming method than questionnaire-based designs, but generally, observational methods are the most appropriate way to describe and measure processes. It avoids the problems of subjectivity and recall bias inherent in questionnaire-based designs,<sup>62 63</sup> especially in stressful situations. While questionnaire-based assessment provides a more general picture of team members' perceptions of team processes, observation methods capture the actions actually performed by the team members.

Moreover, to assure a valid assessment of team processes, observation systems should be as holistic and detailed as possible instead of focusing on a single behavioural facet. The observation system should allow for categorizing all behaviours performed by the team to investigate the interactions between different team behaviours and their relative contribution to the outcome.

The two observation systems used most frequently in the selected studies was the observation method of Manser and colleagues<sup>53</sup> and the behavioural marker system Surgical NOTECHS.<sup>44</sup> The system of Manser and colleagues assesses different aspects of team coordination including information management, task management, coordination via work environment, and others. The NOTECHS system includes behavioural dimensions such as leadership, teamwork, problem-solving, situation awareness, etc. The difference between these two systems is that the former is descriptive, that is, it objectively records actions of the team continuously without any evaluation. Other authors also use descriptive, non-evaluative systems.<sup>31 38 43 52</sup> In the Surgical NOTECHS system, the target behaviours are rated on a scale from 1 to 4 for a defined teamwork episode (e.g. anaesthesia induction). This evaluative component may artificially increase the relationship with performance ratings, while descriptive observation systems provide more objective data on the team process. Thus, it is critical to define performance measures that are truly independent of the team process measures.

The ultimate outcome of high performance in healthcare should be patient safety. As patient safety itself is difficult to measure and to relate to specific team process, various proxy measures have been used. The studies included in this review used many different measures to assess clinical

performance that can be grouped into process performance measures and outcome performance measures.

Outcome performance measures are related to the result of the actions and depend on more than just individuals' behaviour.<sup>21</sup> For example, it is known from resuscitation that the duration of a patient's arrest, the primary arrhythmia, and patient age are better predictors for survival than the actual performance of the clinicians performing the resuscitation.<sup>64</sup> In clinical settings, it is impossible to take in to account all the factors potentially influencing performance, but there are ways to control some of them. For example, the ASA patient classification index has been used to classify patients' risk for complications taking into account the history of the patient<sup>47</sup> and the score for neonatal acute physiology (SNAP) has been used to assess the possibility of complications accounting for the newborn's physiology.<sup>24</sup> Another way to control or balance for confounders are large sample sizes that are often not feasible for very detailed, resource-intensive analyses of team processes and sometimes difficult to obtain in healthcare; especially in field studies requiring a high number of specific, comparable cases performed by care providers with predefined experience levels. In addition, ethical issues sometimes limit the spectrum of cases that can be studied using live observation in clinical settings.

Besides outcome performance measures, the processes leading to this outcome are also good indicators for performance (e.g. timely start of the correct treatment for the patient). These process performance measures refer to what an individual does in a specific work situation and are therefore less influenced by other factors.<sup>21</sup> Process performance can be assessed in almost every setting. During simulation, where it is hardly possible to assess patient-related outcomes, process performance measures are preferable. Performance checklists, for example, that take into account the most important actions for a specific treatment and evaluate those across the whole process provide a valid and reliable method to assess process performance if developed systematically. This includes a theoretical foundation and an integration of official guidelines and experiences of several experts<sup>65</sup> (e.g. through a Delphi process as, for example, done by Burtscher and colleagues).<sup>39 66</sup>

For intervention studies, the results of our review showed that training targeting team process behaviours do influence various outcomes. All the interventions focused exclusively on outcome performance measures. Therefore, one can only assume that the interventions influenced the team processes, which in turn led to better outcomes. This assumption will require further empirical testing to improve our understanding of the mechanisms through which the improvements have been achieved. Unfortunately, no effect sizes could be calculated for most studies, so it is difficult to determine how strong these effects really are. Also, each study referred to a different intervention, none of which was sufficiently described to be reproducible (for a discussion of this issue, see also Buljac-Samadzic and colleagues).<sup>67</sup>

## Limitations

Several limitations of this systematic review have to be taken into account when interpreting the results. We focused only on English, peer-reviewed articles and did not include books or grey literature, so we may have missed relevant publications. Owing to the difficulties with publishing non-significant results,<sup>68</sup> there may be other studies which found no effect of team process behaviour or interventions on performance which we could not access.

In this review, we listed the team processes as they appear in the selected articles. However, if two studies used the same term, this does not necessarily mean they also referred to the same definition of this team process. Furthermore, we focused exclusively on the relationship between team processes and outputs. However, we acknowledge that team processes are not independent of input factors. Specific input factors could neutralize the relationship between processes and outputs. For example, Burtscher and colleagues<sup>50</sup> found a relationship between team monitoring behaviour and performance only when the team members had a shared mental model of the task.

## Future research

This review identified some gaps in the literature on the relationship between team process behaviours and clinical performance.

Since most studies focus on acute patient care, more research needs to be done in other domains of healthcare such as long-term care. Also, only two studies included in this review conducted a follow-up post-test to check if the interventions also had a long-term effect. Thus, studies investigating team processes using a longitudinal design are needed; especially for intervention studies.

In comparison with the sizable literature on the importance of team process behaviour in healthcare, little research has actually investigated the statistical effects on process or outcome performance. To achieve this, valid process performance measures are required and will have to be developed systematically. That is, the relationship of process performance (e.g. checklist-based assessments) and outcomes has to be tested in controlled clinical studies to assure their validity and reliability for assessing performance in clinical and simulated settings.

Of course, there is no single best performance measure. In occupational psychology, it is widely accepted that performance is a multidimensional construct.<sup>15 21</sup> Thus, to get an accurate picture of performance, future studies should use multiple process performance measures or even combinations of process and outcome performance measures.<sup>24 37 44 45</sup>

To further our understanding of specific team processes such as coordination or leadership studies using the same observation systems and performance measures are needed. We gave a brief overview including pros and cons of different measurement methods and future research should take these considerations into account. This will result in more conceptual and methodological consistency and more definitive

findings about the effects of team process behaviours on performance (e.g. supported by meta-analyses).

Our results suggest that team processes in general are clinically relevant because they have an effect on patient outcomes. A large effect size is an indicator for high clinical relevance; however, they are not necessarily linked.<sup>69</sup> For a more precise assessment of clinical relevance, future research should include other factors than statistical results as well.

Some studies included in this review show rather small or no correlations between team processes and performance.<sup>48 49</sup> It is not certain if this is due to unclear or inconsistent definitions of the constructs, validity issues, or confounders. However, we are sure that future research will help to explain and clarify these contradictory results with (i) clear and consistent definitions of the team processes investigated and (ii) more complete descriptions of the mechanisms linking specific team processes to specific performance measures that is embedded in a theoretical framework. Lingard and colleagues<sup>70</sup> illustrate how this could be done using the example of communication patterns related to collaborative work processes and patient safety. In this way, future research will deliver a more accurate picture of the relationship between team processes and performance. With this knowledge, we will be able to design more effective and successful team interventions and implementation strategies which will help to improve patient safety.

## Supplementary material

Supplementary material is available at *British Journal of Anaesthesia* online.

## Acknowledgements

We are grateful to Mariel Dardel for screening the literature and her help in the selection process of relevant articles.

## Declaration of interest

None declared.

## Funding

This work was supported by the Swiss National Science Foundation (grant number, PP00P1\_128616).

## References

- 1 Manser T. Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. *Acta Anaesthesiol Scand* 2009; **53**: 143–51
- 2 Manojlovich M, DeCicco B. Healthy work environments, nurse-physician communication, and patients outcomes. *Am J Crit Care* 2007; **16**: 536–43
- 3 Kohn LT, Corrigan J, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press, 1999
- 4 Salas E, Rosen MA, King H. Managing teams managing crises: principles of teamwork to improve patient safety in the emergency room and beyond. *Theor Issues Ergon Sci* 2007; **8**: 381–94

- 5 Reader TW, Flin R, Mearns K, Cuthbertson BH. Developing a team performance framework for the intensive care unit. *Crit Care Med* 2009; **37**: 1787–93
- 6 Martin JS, Ummenhofer W, Manser T, Spirig R. Interprofessional collaboration among nurses and physicians: making a difference in patient outcome. *Swiss Med Wkly* 2010; **140**: w13062
- 7 Künzle B, Kolbe M, Grote G. Ensuring patient safety through effective leadership behaviour: a literature review. *Safety Sci* 2010; **48**: 1–17
- 8 Rutherford J, Flin R, Mitchell L. Non-technical skills of anaesthetic assistants in the perioperative period: a literature review. *Br J Anaesth* 2012; **109**: 27–31
- 9 McGrath JE. *Social Psychology: A Brief Introduction*. Holt, New York: Rinehart and Winston, 1964
- 10 Gladstein DL. Groups in context: a model of task group effectiveness. *Adm Sci Quart* 1984; **29**: 499–517
- 11 Ilgen DR, Hollenbeck JR, Johnson M, Jundt D. Teams in organizations: from input–process–output models to IMOI models. *Annu Rev Psychol* 2005; **56**: 517–43
- 12 Hackman JR, Morris CG. Group tasks, group interaction process, and group performance effectiveness: a review and proposed integration. In: Berkowitz L, ed. *Advances in Experimental Social Psychology*. New York: Academic Press, 1975; 45–99
- 13 Hackman JR. The design of work teams. In: Lorsch JW, ed. *Handbook of Organizational Behavior*. Englewood Cliffs, NJ: Prentice-Hall, 1987; 315–42
- 14 Jarboe S. A comparison of input–output, process–output, and input–process–output models of small group problem-solving effectiveness. *Commun Monogr* 1988; **55**: 121–42
- 15 Tschan F, Semmer N, Vetterli M, Gurtner A, Hunziker S, Marsch SU. Developing observational categories for group process research based on task and coordination requirement analysis: examples from research on medical emergency-driven teams. In: Boos M, Kolbe M, Kappeler P, Ellwart T, eds. *Coordination in Human and Primate Groups*. Berlin: Springer, 2011; 93–115
- 16 Mathieu JE, Heffner TS, Goodwin GF, Salas E, Cannon-Bowers JA. The influence of shared mental models on team process and performance. *J Appl Psychol* 2000; **85**: 273–83
- 17 Flin R. Identifying and training non-technical skills for teams in acute medicine. *Qual Saf Health Care* 2004; **13**: i80–4
- 18 Champion MA, Medsker GJ, Higgs AC. Relations between work group characteristics and effectiveness: implications for designing effective work groups. *Person Psychol* 1993; **46**: 823–50
- 19 Marks MA, Mathieu JE, Zaccaro S. A temporally based framework and taxonomy of team processes. *Acad Manage Rev* 2001; **26**: 356–76
- 20 Anderson N, Ones DS, Sinangil HK, Viswesvaran C. *Handbook of Industrial, Work and Organizational Psychology: Personnel Psychology*. London: Sage Publications Ltd, 2001
- 21 Sonnentag S, Frese M. Performance concepts and performance theory. In: Sonnentag S, ed. *Psychological Management of Individual Performance*. Chichester: John Wiley & Sons, Ltd, 2002; 1–25
- 22 Nerdinger FW, Blickle G, Schaper N. *Arbeits-und Organisationspsychologie*. Berlin: Springer, 2011
- 23 Davenport D, Henderson W, Mosca C, Khuri S, Mentzerjr R. Risk-adjusted morbidity in teaching hospitals correlates with reported levels of communication and collaboration on surgical teams but not with scale measures of teamwork climate, safety climate, or working conditions. *J Am Coll Surg* 2007; **205**: 778–84
- 24 Pollack MM, Koch MA. Association of outcomes with organizational characteristics of neonatal intensive care units. *Crit Care Med* 2003; **31**: 1620–9
- 25 Adler MD, Vozenilek JA, Trainor JL, et al. Comparison of checklist and anchored global rating instruments for performance rating of simulated pediatric emergencies. *Simul Healthc* 2010; **6**: 18–24
- 26 Donoghue A, Nishisakia A, Suttona R, Halesc R, Boulet JR. Reliability and validity of a scoring instrument for clinical performance during Pediatric Advanced Life Support simulation scenarios. *Resuscitation* 2010; **81**: 331–6
- 27 Wiles R, Robison J. Teamwork in primary care: the views and experiences of nurses, midwives and health visitors. *J Adv Nurs* 1994; **20**: 324–30
- 28 Körner M, Bengel J. Teamarbeit und Teamerfolg bei multi-und interdisziplinären Teams in der medizinischen Rehabilitation teamwork and team success in multidisciplinary and interdisciplinary teams in medical rehabilitation. *Rehabilitation* 2004; **43**: 348–57
- 29 Mishra A, Catchpole K, Dale T, McCulloch P. The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. *Surg Endosc* 2007; **22**: 68–73
- 30 Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009; **6**: e1000100
- 31 Tschan F, Semmer NK, Gautschi D, Hunziker P, Spychiger M, Marsch SU. Leading to recovery: group performance and coordinative activities in medical emergency driven groups. *Hum Perform* 2006; **19**: 277–304
- 32 Thomas EJ. Improving teamwork in healthcare: current approaches and the path forward. *Br Med J* 2011; **20**: 647–50
- 33 Buckley S, Coleman J, Davison I, et al. The educational effects of portfolios on undergraduate student learning: a Best Evidence Medical Education (BEME) systematic review. BEME Guide No. 11. *Med Teach* 2009; **31**: 282–98
- 34 Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998; **52**: 377–84
- 35 Kirk RE. Promoting good statistical practices: some suggestions. *Educ Psychol Meas* 2001; **61**: 213–8
- 36 ElBardissi AW, Wiegmann DA, Henrickson S, Wadhera R, Sundt Iii TM. Identifying methods to improve heart surgery: an operative approach and strategy for implementation on an organizational level. *Eur J Cardiothorac Surg* 2008; **34**: 1027–33
- 37 Manojlovich M, Antonakos CL, Ronis DL. Intensive care units, communication between nurses and physicians, and patients' outcomes. *Am J Crit Care* 2009; **18**: 21–30
- 38 Tschan F, Semmer NK, Gurtner A, et al. Explicit reasoning, confirmation bias, and illusory transactive memory: a simulation study of group medical decision making. *Small Group Res* 2009; **40**: 271–300
- 39 Burtcher MJ, Wacker J, Grote G, Manser T. Managing nonroutine events in anesthesia: the role of adaptive coordination. *Hum Fact* 2010; **52**: 282–94
- 40 Burtcher MJ, Manser T, Kolbe M, et al. Adaptation in anaesthesia team coordination in response to a simulated critical event and its relationship to clinical performance. *Br J Anaesth* 2011; **106**: 1–6
- 41 Manser T, Harrison TK, Gaba DM, Howard SK. Coordination patterns related to high clinical performance in a simulated anesthetic crisis. *Anesth Analg* 2009; **108**: 1606–15
- 42 Carlson J, Min E, Bridges D. The impact of leadership and team behavior on standard of care delivered during human patient

- simulation: a pilot study for undergraduate medical students. *Teach Learn Med* 2009; **21**: 24–32
- 43 Marsch SCU, Müller C, Marquardt K, Conrad G, Tschan F, Hunziker PR. Human factors affect the quality of cardiopulmonary resuscitation in simulated cardiac arrests. *Resuscitation* 2004; **60**: 51–6
- 44 Catchpole KR, Giddings AE, Wilkinson M, Hirst G, Dale T, de Leval MR. Improving patient safety by identifying latent failures in successful operations. *Surgery* 2007; **142**: 102–10
- 45 Catchpole K, Mishra A, Handa A, McCulloch P. Teamwork and error in the operating room: analysis of skills and roles. *Ann Surg* 2008; **247**: 699–706
- 46 Schraagen JM, Schouten T, Smit M, et al. A prospective study of paediatric cardiac surgical microsystems: assessing the relationships between non-routine events, teamwork and patient outcomes. *Br Med J* 2011; **20**: 599–603
- 47 Mazzocco K, Petitti DB, Fong KT, et al. Surgical team behaviors and patient outcomes. *Am J Surg* 2009; **197**: 678–85
- 48 Thomas E, Sexton J, Lasky R, Helmreich R, Crandell D, Tyson J. Teamwork and quality during neonatal care in the delivery room. *J Perinatal* 2006; **26**: 163–9
- 49 Westli H, Johnsen B, Eid J, Rasten I, Brattebø G. Teamwork skills, shared mental models, and performance in simulated trauma teams: an independent group design. *Scand J Trauma Resusc Emerg Med* 2010; **18**: 47
- 50 Burtcher MJ, Kolbe M, Wacker J, Manser T. Interactions of team mental models and monitoring behaviors predict team performance in simulated anesthesia inductions. *J Exp Psychol Appl* 2011; **17**: 257
- 51 Siassakos D, Fox R, Crofts JF, Hunt LP, Winter C, Draycott TJ. The management of a simulated emergency: better teamwork, better performance. *Resuscitation* 2011; **82**: 203–6
- 52 Künzle B, Zala-Mezo E, Kolbe M, Wacker J, Grote G. Substitutes for leadership in anaesthesia teams and their impact on leadership effectiveness. *Eur J Work Organ Psychol* 2010; **19**: 505–31
- 53 Manser T, Howard SK, Gaba DM. Adaptive coordination in cardiac anaesthesia: a study of situational changes in coordination patterns using a new observation system. *Ergonomics* 2008; **51**: 1153–78
- 54 Kalisch BJ, Curley M, Stefanov S. An intervention to enhance nursing staff teamwork and engagement. *J Nurs Adm* 2007; **37**: 77–84
- 55 McCulloch P, Mishra A, Handa A, Dale T, Hirst G, Catchpole K. The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre. *Qual Saf Health Care* 2009; **18**: 109–15
- 56 Wolf FA, Way LW, Stewart L. The efficacy of medical team training: improved team performance and decreased operating room delays: a detailed analysis of 4863 cases. *Ann Surg* 2010; **252**: 477–83
- 57 Morey JC, Simon R, Jay GD, et al. Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MedTeams project. *Health Serv Res* 2002; **37**: 1553–81
- 58 Nielsen PE, Goldman MB, Mann S, et al. Effects of teamwork training on adverse outcomes and process of care in labor and delivery: a randomized controlled trial. *Obstet Gynecol* 2007; **109**: 48–55
- 59 Phipps MG, Lindquist DG, McConaughy E, O'Brien JA, Raker CA, Paglia MJ. Outcomes from a labor and delivery team training program with simulation component. *Am J Obstet Gynecol* 2011; **206**: 1–7
- 60 Fernandez Castelao E, Russo SG, Cremer S, et al. Positive impact of crisis resource management training on no-flow time and team member verbalisations during simulated cardiopulmonary resuscitation: a randomised controlled trial. *Resuscitation* 2011; **82**: 1338–43
- 61 Burke CS, Stagl KC, Salas E, Pierce L, Kendall D. Understanding team adaptation: a conceptual analysis and model. *J Appl Psychol* 2006; **91**: 1189–207
- 62 Bortz J, Döring N. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer Verlag, 2006
- 63 Greve W, Wentura D. *Wissenschaftliche Beobachtung: Eine Einführung*. Weinheim: BeltzPVU, 1997
- 64 Cooper S, Cade J. Predicting survival, in-hospital cardiac arrests: resuscitation survival variables and training effectiveness. *Resuscitation* 1997; **35**: 17–22
- 65 Landon BE, Normand SLT, Blumenthal D, Daley J. Physician clinical performance assessment. *J Am Med Assoc* 2003; **290**: 1183–9
- 66 Gordon TJ. *The Delphi Method*. In: Glenn JC, ed. Futures Research methodology series. Washington, D. C.: United Nations University, 1994
- 67 Buljac-Samarzic M, Dekker-van Doorn CM, Van Wijngaarden JDH, Van Wijk KP. Interventions to improve team effectiveness: a systematic review. *Health Policy* 2010; **94**: 183–95
- 68 Easterbrook PJ, Gopalan R, Berlin J, Matthews DR. Publication bias in clinical research. *Lancet* 1991; **337**: 867–72
- 69 Ogles BM, Lunnen KM, Bonesteel K. Clinical significance: history, application, and current practice. *Clin Psychol Rev* 2001; **21**: 421–46
- 70 Lingard L, Whyte S, Espin S, Ross Baker G, Orser B, Doran D. Towards safer interprofessional communication: constructing a model of 'utility' from preoperative team briefings. *J Interprof Care* 2006; **20**: 471–83
- 71 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum, 1988
- 72 Schraagen JM. Dealing with unforeseen complexity in the OR: the role of heedful interrelating in medical teams. *Theor Issues Ergon Sci* 2011; **12**: 256–72
- 73 Rasch B, Frieze M, Hofmann W, Naumann E. *Quantitative Methoden: Einführung in die Statistik 1*. 2 Edn. Heidelberg: Springer Medizin Verlag, 2006
- 74 Walker DA. JMASM9: converting Kendall's tau for correlational or meta-analytic analyses. *J Mod Appl Stat Meth* 2003; **2**: 525–30

Handling editor: R. P. Mahajan