



Assessing and improving teamwork in cardiac surgery

Jan Maarten Schraagen, Ton Schouten, Meike Smit, et al.

Qual Saf Health Care 2010 19: 1-6
doi: 10.1136/qshc.2009.040105

Updated information and services can be found at:
<http://qualitysafety.bmj.com/content/19/6/1.59.full.html>

These include:

References

This article cites 26 articles, 13 of which can be accessed free at:
<http://qualitysafety.bmj.com/content/19/6/1.59.full.html#ref-list-1>

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:
<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:
<http://journals.bmj.com/cgi/ep>

Assessing and improving teamwork in cardiac surgery

Jan Maarten Schraagen,¹ Ton Schouten,³ Meike Smit,⁴ Felix Haas,⁵
Dolf van der Beek,⁴ Josine van de Ven,¹ Paul Barach^{2,3}

¹TNO Human Factors, Soesterberg, The Netherlands

²Patient Safety Center of the University Medical Center Utrecht, Utrecht, The Netherlands

³Department of Perioperative Care and Emergency Medicine, University Medical Center Utrecht, Utrecht, The Netherlands

⁴TNO Quality of Life, Leiden/Hoofddorp, The Netherlands

⁵Department of Paediatric Cardiothoracic Surgery, University Medical Center Utrecht, Utrecht, The Netherlands

Correspondence to

Professor Jan Maarten Schraagen, TNO Human Factors, PO Box 23, 3769 ZG Soesterberg, The Netherlands; jan_maarten.schraagen@tno.nl

Accepted 1 May 2010

ABSTRACT

Objective Cardiac surgery (PCS) has a low error tolerance, is dependent upon sophisticated organisational structures and demands high levels of cognitive and technical performance. The aim of the study was to assess the role of intraoperative non-routine events (NREs) and team performance on paediatric cardiac surgery outcomes. The current paper focuses on improving methods for studying teamwork; a companion paper will report on the empirical results.

Methods The authors trained human factors observers to observe and code the NRE's and teamwork from time of arrival of the patient into the operating room (OR) to the patient handover in the intensive care unit. The observers underwent immersive training in which each observer attended 10 operations, learnt in detail about the technical procedures and clinical tasks and received practice in coding teamwork. Two observers were used interchangeably to observe OR teamwork. The authors instigated a rigorous training and assessment protocol, with independent assessment of their performance by both senior medical and human factors experts using video-based assessment. Real-time teamwork observations were supplemented with process mapping, questionnaires on safety culture, level of preparedness by the team, difficulty of the operation and outcome measures.

Results 19 PCS cases were observed. The observers observed a total of 255 hr of operations, including the first 10 training cases. We found that 68% of events were observed by one observer but only 32% of all events were observed by both observers. If an event was coded by both observers, 76% was coded in the same way, resulting in high levels of inter-rater agreement. The inter rater reliability of the four main teamwork categories was 91% with Cohen kappa of 0.77. Recommendations were developed for observing teamwork in the operating room, for instance 'train observers on video recordings of real operations (not scripted performance), preferably of at least 1–2 h in duration' and 'Rate teamwork in real time and not afterwards.'

Conclusions PCS is an ideal model to explore team performance. A challenge for the future is to make observations of teamwork in healthcare settings more efficient and robust.

One of the riskiest and most complex in-hospital environments is the paediatric cardiac surgical (PCS) suite.¹ PCS procedures are serious paediatric interventions, involving complex, frequently critically ill patients, with anatomical diversity, haemodynamic vulnerability, and the need for a highly skilled, multispecialty team.² PCS is a specialty with low error tolerance, encompasses many complex proce-

dures that are dependent upon sophisticated organisational structures, requires coordinated efforts of multiple individuals and demands high levels of cognitive and technical performance.³

Several factors have been linked to poor outcomes in PCS, including institutional and surgeon-specific volumes,⁴ complexity of cases⁵ and systems failures.¹ The importance of human factors and systems research in improving outcomes for paediatric cardiac surgery has been highlighted in the Bristol Royal Infirmary⁷ and the Manitoba Inquiries.⁸ However, after a remarkable decrease in adverse outcomes over the last two decades, interventions and change strategies have had a limited impact on improving PCS outcomes further. This may be attributable to a lack of appreciation of the evidence about human factors in PCS, including a poor understanding of the complexity of interactions between the technical task, stressful PCS setting, rigid staff hierarchies, lack of time to brief and debrief, and resistance to change.⁶

One of the contributing factors to non-routine events is defective teamwork.⁹ Breakdowns in teamwork in the operating room may lead to errors¹⁰ and poor outcomes.¹¹ Technical skills are fundamental to good outcomes but non-technical factors also impact significantly on individual and team performance and patient outcomes.¹² Observing and assessing teamwork in situ is therefore desirable to derive interventions to overcome deficiencies in teamwork and improve outcomes.

The aim of this study was to assess the role of intraoperative non-routine events (NREs) and team performance on outcomes during paediatric cardiac surgery. Human-factors observers were trained to observe NREs and score teamwork in real time on a second-to-second basis. We employed a multilevel measurement approach by also incorporating input factors such as the process of care, complexity of the operation and the team members' level of preparedness, as well as the outcome of the operation. Our goal was to establish a rigorous, evidence-based approach to quality improvement, by doing a prospective pre- and postintervention study.¹³ This article reports on the qualitative aspects of the preintervention observations, focusing on improving methods for observing teamwork. A companion article will focus on the quantitative results of the full pre–post test study.

METHODS

We studied paediatric cardiac surgery at the Wilhelmina Children's Hospital, part of the University Medical Center Utrecht in The Netherlands. Two observers performed real-time

Original research

prospective observations of the PCS team from the inception of anaesthesia to the patient handover in the intensive care unit.¹⁴ The study included detailed process mapping, a comprehensive cognitive task analysis, detailing of the PCS team¹⁵ and training the observers in a validated and reliable manner all described elsewhere.⁶ Full institutional review board approval was attained. Written consent was acquired from all PCS team members.

Clinical case complexity was measured using the comprehensive Aristotle risk assessment scoring system.¹⁶ This scoring tool stratifies based on the potential for morbidity, mortality and the anticipated technical difficulty of a given procedure.

We used Weinger and Slagle's¹⁷ definition of 'non-routine event' adapted from the nuclear power industry, namely: 'any event that is perceived by care providers or skilled observers to be unusual, out-of-the-ordinary or atypical.' This is a broad definition and includes everything from phone calls, masks not worn properly to serious incidents endangering the patient's condition. We further subdivided these non-routine events into 'task-related NREs' and 'non-task related NREs.' Task-related NREs are related to the task of team members as a technical skill. Non-task related NREs are by definition all the other NREs, unrelated to the task to perform as a technical skill. Non-routine events were assessed by the observers during the operation, and corroborated by the team members afterwards, both by asking about non-routine events in a questionnaire and by interviewing team members using an open-ended validated tool.

Levels of preparedness were assessed by a brief four-question questionnaire that was administered to each team member beforehand. This questionnaire was based on an earlier validated questionnaire about physical and mental fitness, staffing level and concerns about equipment status.¹⁸ Team member knowledge, skills and attitudes were partly captured in the process checklist and the cognitive task analysis, but not detailed to the level of each individual team member.

The surgical outcome was determined afterwards in terms of three categories:¹⁹ uncomplicated, minor complications, major morbidity.

Figure 1 presents our adaptation of the well-known Conditions–Processes–Outcomes model^{20–21} of teamwork, showing the relationships among the various variables described above. The conditions mentioned are illustrative and not meant to be exhaustive.

Selection of coding scheme

Measurement needs to assess process as well as outcome.²² Although outcome measures seem deceptively simple, appear to be objective and are easy to obtain, they are insufficient in gauging the process of care and have a number of drawbacks when they are the only assessment tool used. First, a good outcome does not necessarily suggest optimal processes or that effective teamwork was executed.²³ Good outcomes may come about by sheer luck or patient characteristics. Second, outcome measures provide little guidance on 'why' something happened, and they are therefore not particularly useful for providing practitioners or trainees with feedback. Third, good outcomes may actually be the result of poor processes. When the only feedback provided to such a team is outcome-based, these unclear/flawed processes may be inadvertently reinforced.²⁴ Process measures provide the information needed for diagnostic feedback that allows teams to decide the exact areas in need of remediation.

Classification and coding schemes need to assess team processes, such as communication and team leadership (collectively referred to as 'teamwork'). Team process measures rely almost exclusively

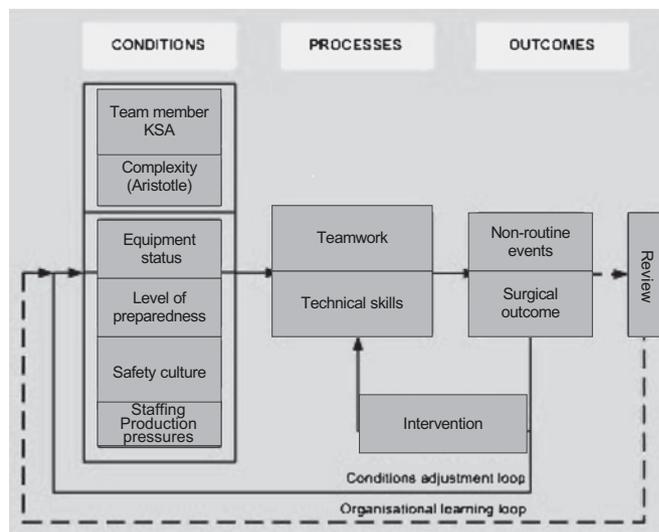


Figure 1 Team performance framework. KSA, knowledge, skills, attitudes.

on observation.²⁵ One lesson learnt with observations of military command and control teams is that observations should focus on observable behaviours rather than internal states of mind.²⁴ This is why the construct of Situation Awareness is frequently found to be less reliable than other constructs that are more tied to observable behaviour (but see Wright and Endsley²⁶ for alternatives). A second lesson learnt is that approximately three to five dimensions are sufficient to adequately describe the quality and robustness of the observed teamwork.²⁴

There are several teamwork classification tools available, some specific for surgeons and anaesthetists (Non-technical Skills for Surgeons (NOTSS), Non-technical skills (NOTECHS), Observational Teamwork Assessment for Surgery (OTAS[®]), Anaesthetists Non-Technical Skills (ANTS); see below), others being more general.²⁷ Yule *et al* reported a study in which the reliability of the NOTSS tool was evaluated.²⁸ The authors note that high levels of sensitivity and inter-rater reliability are not achieved unless raters have proper training and calibration. They recommend at least 2 days training for using this type of rating system. We subjected our observers to 10 operations over several weeks for training purposes, as well as several sessions of coding real videotaped behaviour where discrepancies in coding were discussed and settled upon using a 'gold' standard.

Fletcher *et al* have reported on the evaluation of the ANTS behavioural marker system.²⁹ The results showed inter-rater agreement levels of between 0.55 and 0.67. This is not as high as would be accepted in other industries but, given the limited training provided, was nevertheless deemed acceptable. The limitations of this study, as with the NOTSS study discussed above, are the use of scripted scenarios rather than real-life situations, and the limited time available for training. We have trained our observers on videorecordings of real operations, not scripted behaviour.

We classified teamwork aspects using four main categories: leadership, situation awareness, decision-making, and teamwork and cooperation (see table 1). The coding scheme was used to rate the effect of the behaviour (present or absent) on teamwork. In particular, for coding the teamwork aspects, we have drawn upon and slightly modified (because of the multidisciplinary nature of the team) the NOTECHS system,³⁰ and the associated ANTS system for coding anaesthetists' non-technical skills²⁹ and the NOTSS system for coding surgeons' non-technical skills.^{31–32} Although our coding scheme shows similarities with

Table 1 Teamwork classification tool and rating scale

Leadership	Maintenance of standards Workload management Support of others
Situation awareness	Notice Understand Think ahead
Decision-making	Definition and diagnosis Risk assessment and option generation Outcome review
Teamwork and cooperation	(Relational) communication Feedback Backup behaviour Coordination
6	The demonstrated behaviour was highly effective in enhancing teamwork
5	High level of enhancement of teamwork through demonstrated behaviour
4	Moderate enhancement of teamwork through demonstrated behaviour
3	Behaviour neither hindered nor enhanced teamwork
2	Slight detriment of teamwork through lack of certain essential behaviour
1	Teamwork was compromised through lack of certain essential behaviour
0	The demonstrated behaviour was unclear and hindered teamwork

the Communication and Teamwork Skills (CATS) Assessment,³³ we have placed more emphasis on the behavioural markers of 'maintenance of standards,' 'risk assessment and option generation'. We have used a broader interpretation of the concept of Situation Awareness to include all communications about the real-time status of the patient or the actions carried out by the team members (eg, 'Ventilation start'; 'ACT 150'; 'Cardioplegia stopped at 45 ml'). The description of the seven-point rating scales themselves was derived from the OTAS[®] research instrument.³⁴ Behaviour was rated along one dimension only, namely whether it hindered or enhanced teamwork. Using one dimension to rate behaviour is conceptually easier for observers and less ambiguous than the two dimensions (patient safety and teamwork) employed simultaneously by the NOTECHS system. Teamwork aspects were included in a detailed process checklist, consisting of a detailed task analysis of 15 phases involved in the surgical process.³⁵ Hence, for each subtask, the occurrence of certain activities, the occurrence of non-routine events, as well as the relevant teamwork relating to these adverse events was recorded and rated. All communications among team members were noted in real time and written down on a scoring form. These communications were classified and rated during the operation, rather than afterwards, in order to avoid hindsight bias. At this stage, there was no aggregation of the ratings. Teamwork was classified and rated separately for each discipline (surgeons, anaesthetists, perfusionists, nurses). This resulted in dozens of teamwork scores for each discipline for each operation. Aggregation of the teamwork scores was done afterwards by discipline, the teamwork (sub)category or a combination of the two. Non-routine events were scored separately from the teamwork, as events (such as beepers going off) and teamwork (such as dealing with these beepers) are conceptually different.

Selection of observers

It should be noted that our observers were human factors experts rather than medical professionals. This allowed the

study to maintain a detached and objective view on the events involved. The non-routine events observed were afterwards validated by these medical professional as medical professionals provide context, valuation, relevance and clarity to a human factors partner.³⁶ Although we acknowledge that not all non-routine events are task-related, it is critical to obtain a complete overview of all process variations, as these events could accumulate and add up to form a major event.^{6,9}

It should also be noted that we were not primarily interested in the level of the technical skills of surgeons or anaesthetists, so extensive domain knowledge was not required on the part of our observers. Rather, they needed to be sufficiently familiar with the technical steps of the domain to be able to note and document variations in the process that were defined as non-routine events. Admittedly, they were not able to detect specialised surgical mishaps, unless these were mentioned by the clinical team members themselves during the postsurgical interviews or in the questionnaires.

Training of observers

Baker *et al* have summarised the available performance appraisal literature up to 1999.³⁷ They concluded that raters should be trained to evaluate non-technical skills using expert standards, so they will adopt a common frame of reference. This implies that raters should receive feedback that compares their ratings to standards that are established by experts. Raters should be trained on videotapes that display actual performance, as opposed to scripted performance, because 'actual performance typically contains more subtle variations that are harder for raters to observe and distinguish.'³⁷ In our study, a human factors teamwork expert and an expert cardiac anaesthesiologist both watched the videotapes and provided feedback to the raters helping to calibrate their observations.

In our study, training for the two observers included in-depth directed study of cardiac surgery theory and literature, watching videotaped paediatric cardiac surgery procedures, and detailed discussions of ethnographic observational methods.^{6,12} The observers were encouraged to ask questions, and were informally tested by paediatric cardiac surgeons, cardiac anaesthesiologists, cardiac perfusionists and OR nurses to ascertain the observer's knowledge and understanding of procedures and operating room dynamics and culture. Observers observed at least 10 live cases over several weeks prior to collecting data. Prior to collecting data, observers had to pass an examination. The exam involved watching a 2 h fragment of a videotaped operation and scoring this fragment in real time on a second-by-second basis. Inter-rater reliability was assessed by calculating the number of events scored by both observers and determining whether or not observers rated these events identically as far as teamwork was concerned. The inter-rater reliability was 91% at the level of the four main teamwork categories and 84% at the level of the 14 detailed subcategories. Taking chance into account, the Cohen κ was 0.77, which shows a high level of agreement.³⁸

In our study, acquisition of relevant domain knowledge by the observers was assessed by an experienced anaesthetist, in the following ways. First, the anaesthetist interviewed the observers after the exam to check whether they had understood the procedure and whether they had noted the non-routine events involved. Second, the anaesthetist developed a set of 23 statements that could be either right or wrong. The set of statements was administered during a second exam midway through the first observation period. These statements were geared at the level of an experienced anaesthesia resident trained in paediatric cardiac anaesthesia. Taking into account chance levels, observers

Original research

scored a '5' on a scale from 0 to 10. The level of knowledge gained by the observers of the procedures, although non-trained anaesthetists was akin to the level of a third year resident and the score was considered acceptable. Moreover, both observers scored equally well.

RESULTS

Inter-rater reliability

In the current study, inter-rater reliability assessment was carried out on all categories coded by the observers. For this purpose, segments of video lasting 1–2 h were taken from actual operations and were used for assessing inter-rater reliability. The results showed, first, that 68% of the observations were unique, in the sense that of all the events noted by the observers, 68% were observed by only one observer. Second, of those events that were recorded identically, 65% were coded by both observers. Some events, although recorded, were not coded as being teamwork, or were overlooked. Third, of the events that were coded by both observers, 76% were coded in the same way. From this latter figure, we may conclude that the coding scheme itself led to high levels of inter-rater agreement. The results suggest that the subjectivity lies earlier in the observation process, namely in the decision to cognitively attend to a particular event and then to code it. This is understandable, given the multitude of events that goes on simultaneously during a complex PCS operation. Once an event is coded within the teamwork rating system, there is substantial agreement between the observers as to the category to which a particular behaviour belongs.

Hence, out of every 100 events occurring during an operation, only 32 are observed by both observers, 21 of these are coded by both observers, and 16 are coded by both observers in an identical way (see figure 2 for an overall summary).

In our study, we tested the observers again after approximately 10 operations (midway through the preintervention session) on a 1 h video fragment of a real operation. The Cohen κ score in this case was 0.50. Although still acceptable, this score was lower than the initial 0.77 with which the observers had finished their training period. The reason for this is unclear. It may be partly due to the length of the video fragment (1 h instead of 2 h) or to the lack of standard teamwork displayed (this particular 1 h fragment was near the end of the operation, with a relaxed atmosphere, where there may have been fewer opportunities to display teamwork behaviour).

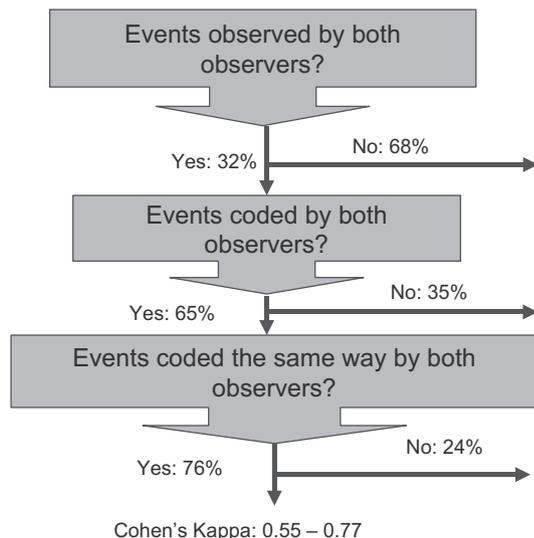


Figure 2 Assessment of inter-rater reliability of the two observers.

Table 2 Examples of non-routine events from one operation

9.30	S1	Pager goes off. N2 calls.
9.35	P2 → A1	Discovers that the NIRS saturation-probes are not attached yet
9.35	A3	Drug syringe pumps are not functioning properly
9.43	S1	Does not ask if the incision can be made but just starts
10.08	S1	During inserting the retractor, S1 cuts his thumb
10.10	P1	Disconnection of needle and syringe during aspiration of antibiotics from a drug vial. Content comes out with high pressure.
10.32	N3	Replaces N2
10.34	P2	Telephone from the pre-operative holding area about the consent form of the patient (is missing)

The first column shows the clock time; the second column shows the clinical team member (S, surgeon; P, perfusionist; A, anaesthetist; N, nurse); the third column shows the non-routine event.

NIRS, near infrared spectroscopy.

In order to assess the stability of our observers' coding capabilities, we also gave them a third 1 h video fragment after 6 months during which they had not observed. The Cohen κ in this case was 0.66, which again indicates substantial agreement. We may therefore conclude that it is possible to train observers, achieve, and sustain high levels of inter-rater reliability, even over long periods of time.

Example of data produced

Two examples of data are provided. First, a selection of a number of NREs is shown, in chronological order (table 2). The first column shows the time, the second column the actor and the third column the NRE.

Second, an extract from the original teamwork scoring system is shown (table 3). This extract only shows teamwork behaviour displayed by the surgeons. The first column shows the time, the second the actors (who communicates with whom), the third the teamwork observed, the fourth the teamwork category and the fifth the rating of the teamwork.

In box 1, we have distilled our main lessons learnt in terms of the methodology of observing teamwork in the OR.²³ We hope these lessons prove to be useful for other researchers interested in improving patient outcomes by coding and observing teamwork.

CONCLUSIONS AND LIMITATIONS

A considerable amount is known about how to observe teamwork, both in healthcare settings and in other settings (military, aviation, aerospace).²² We have reviewed the literature and adapted what is known for our study of assessing the impact of human factors on paediatric cardiac surgical teams. We recognise several limitations to the study. The capture of observational data is by necessity subjective and observer-dependent, and can suffer from inter-rater reliability as well as a sampling bias. Undoubtedly, many events might have been missed.

Table 3 Examples of teamwork actions and their scoring from one operation

10.00	S1 → A1	Is the ACT OK? Yes it's all right.	SA1	3
10.08	S2 → N1	Request a bigger suction device. N1 gives it.	MS	4
10.11	S1 → P1	Is a 12 mm also possible? Yes. Why did I get a 14 mm instead then? That's what you asked for. S1 gets the 12 mm.	WM	4
10.11	S1 → S3	Only do what I tell you to do. Hold your pincet over there.	MS	4
10.15	S1 → P1	Are the pulsations alright? Yes.	SA1	3
10.20	S1 → P1	Cool to 32 degrees.	C	3

The first column shows the clock time; the second column shows the actors (S, surgeon; P, perfusionist; A, anaesthetist; N, nurse); the third column shows the teamwork observed; the fourth column shows the classification code; the fifth column shows the rating (Likert scale ranging from 0 to 6).

ACT, activated clotting time.

Box 1 Recommendations for observing teamwork in the operating room

- ▶ Use a detailed process map to write down observations.
- ▶ Rate both moment-to-moment processes and outcomes.
- ▶ Try out various teamwork classification tools and adjust them to fit the observers' requirements and the context in which observations take place.
- ▶ Use rating scales to judge the quality of teamwork processes; scales should be based on a single dimension (eg, impact on teamwork).
- ▶ Train observers on video recordings of real operations (not scripted performance), preferably 1–2 h in duration.
- ▶ Discuss discrepancies in coding and settle on 'gold standard,' so raters will adopt a common frame of reference.
- ▶ Use video recordings repeatedly to test for inter-rater reliability.
- ▶ Rate teamwork in real time and not afterwards.
- ▶ Verify observations immediately after the operation with personnel involved in the operation (i.e, interviews, questionnaire).
- ▶ Solicit opinions of all team members involved by distributing questionnaires before and after the operation.
- ▶ Observers should remain a 'fly on the wall' during the operation and not become involved in the actual team's work.
- ▶ Sample teamwork over a wide variety of conditions and times.

By objectively assessing the performance of the two observers, we found that the decision to attend to a particular event and to code it may differ from one observer to another. Once a particular behaviour is coded, it is usually coded in the same way by both observers. Given that during the actual operations, for practical reasons (small size of operating room; large number of team members routinely in the room (8–10), and the intense and focused atmosphere of the team) only one observer was present, we may tentatively conclude that the choice of which event to attend to is a subjective one. However, our observational data were corroborated afterwards by questionnaires completed by team members as well as by interviews with team members. Moreover, we assessed the inter-rater reliability of our observers three times in the course of the study and it was acceptable.

We have developed an intervention aimed at identifying NRE's and improving teamwork skills based on the results of this study. We have subsequently assessed the effect of the intervention in a series of 20 postintervention observations, using the same observers we used during the first 19 observations.

A challenge for the future is to make observations of teamwork in healthcare settings more efficient and robust. One option might be to train healthcare professionals who are otherwise not involved in the healthcare team under consideration. However, it remains to be seen whether healthcare professionals are less subjective in their choice of what event to attend to when they are observing clinical procedures.³⁶ Ideally, a combination of a human factors specialist and a healthcare professional may yield the most objective results, as the human factors specialist will not try to explain away protocol violations, and the healthcare professional will not miss patient-related NREs. Training two observers to be both present during an operation is, however, a very time-consuming option. Also, the study results are highly dependent on these scarce and well-trained observers. Yet, without automation of teamwork clas-

sification in sight, the well-trained human observer is the best option we have for understanding and improving teamwork.

Funding TNO Quality of Life, Wassenaarseweg 56, Leiden, The Netherlands.

Competing interests TS and FH are employed by University Medical Center Utrecht and were part of the team under study.

Ethics approval Ethics approval was provided by University Medical Centre Utrecht IRB.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Catchpole KR**, Giddings A, deLeval M, *et al*. Identification of systems failures in successful paediatric cardiac surgery. *Ergonomics* 2006;**49**:567–88.
2. **Mejak B**, Stammers A, Rauch E, *et al*. A retrospective study on perfusion incidents and safety devices. *Perfusion* 2000;**15**:51–61.
3. **Elbardissi A**, Wiegmann D, Dearani J, *et al*. Application of the human factors analysis and classification system methodology to the cardiovascular surgery operating room. *Ann Thorac Surg* 2007;**83**:1412–19.
4. **Hannan E**, Racz M, Kavey R, *et al*. Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. *Pediatrics* 1998;**101**:963–69.
5. **Wong D**, Vander Salm T, Agnihotri A, *et al*. Prospective assessment of intraoperative precursor events during cardiac surgery. *Eur J Cardiothorac Surg* 2006;**29**:447–55.
6. **Barach P**, Johnson JK, Ahmed A, *et al*. A prospective observational study of human factors adverse events, and patient outcomes in surgery for pediatric cardiac disease. *J Thorac Cardiovasc Surg* 2008;**136**:1422–8.
7. **The Bristol Royal Infirmary Inquiry**. *The Inquiry into the Management of Care of Children Receiving Complex Heart Surgery at the Bristol Royal Infirmary. Final Report*. London: The Inquiry, 2001.
8. **Sinclair CM**. *The Report of the Manitoba Pediatric Cardiac Surgery Inquest: an Inquiry into Twelve Deaths at the Winnipeg Health Sciences Centre in 1994*. Winnipeg: Provincial Court of Manitoba, 2000.
9. **De Leval MR**, Carthey J, Wright DJ, *et al*. Human factors and cardiac surgery: a multicenter study. *J Thorac Cardiovasc Surg* 2000;**119**:661–72.
10. **Stevenson KS**, Gibson SC, MacDonald D, *et al*. Measurement of process as quality control in the management of acute surgical emergencies. *Br J Surg* 2006;**94**:376–81.
11. **Christian C**, Gustafon M, Roth E, *et al*. A prospective study of patient safety in the operating room. *Surgery* 2006;**139**:159–73.
12. **Galvan C**, Bacha B, Mohr J, *et al*. A human factors approach to understanding patient safety during pediatric cardiac surgery. *Progress in Pediatric Cardiology* 2005;**20**:13–20.
13. **Mittman BS**. Creating the evidence base for quality improvement collaboratives. *Ann Intern Med* 2004;**140**:897–901.
14. **Sharit J**, McCane L, Thevenin DM, *et al*. Examining links between sign-out reporting during shift changeovers and patient management risks. *Risk Anal* 2008;**28**:983–1001.
15. **Barach P**, Johnson J. Safety by design: understanding the dynamic complexity of redesigning care around the clinical microsystem. *Qual Saf Health Care* 2006;**15**(Suppl 1):i10–16.
16. **Lacour-Gayet F**, Clarke D, Jacobs J, *et al*. The Aristotle score: a complexity-adjusted method to evaluate surgical results. *Eur J Cardiothorac Surg* 2004;**25**:911–24.
17. **Weinger MB**, Slagle J. Human factors research in anesthesia patient safety: Techniques to elucidate factors affecting clinical task performance and decision making. *J Am Med Inform Assoc* 2002;**9**(6 Suppl 1):S58–63.
18. **Bognar A**, Barach P, Johnson J, *et al*. Errors and the burden of errors: attitudes, perceptions and the culture of safety in pediatric cardiac surgical teams. *Ann Thorac Surg* 2008;**85**:1374–81.
19. **Catchpole KR**, Giddings AEB, Wilkinson M, *et al*. Improving patient safety by identifying latent failures in successful operations. *Surgery* 2007;**142**:102–10.
20. **Hackman JR**. The design of work teams. In: Lorsch J, ed. *Handbook of Organizational Behavior*. Englewood Cliffs, NJ: Prentice-Hall, 1987:315–42.
21. **Salas E**, Dickenson TL, Tannenbaum SI, *et al*. Toward an understanding of team performance and training. In: Swezey RW, Salas E, eds. *Teams: Their Training and Performance*. Norwood, NJ: Ablex, 1992:3–29.
22. **Cannon-Bowers JA**, Salas E. A framework for developing team performance measures in training. In: Brannick MT, Salas E, Prince C, eds. *Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates, 1997:45–62.
23. **Salas E**, Burke CS, Fowlkes JE, *et al*. On measuring teamwork skills. In: Thomas JC, ed. *Comprehensive Handbook of Psychological Assessment*. Hoboken, NJ: John Wiley & Sons, 2003:427–42.
24. **Smith-Jentsch KA**, Johnston JH, Payne SC. Measuring team-related expertise in complex environments. In: Cannon-Bowers JA, Salas E, eds. *Making Decisions under Stress: Implications for Individual and Team Training*. Washington, DC: American Psychological Association, 1998:61–87.
25. **Barach P**, Weinger M. Trauma team performance. In: Wilson WC, Grande CM, Hoyt DB, eds. *Trauma: emergency resuscitation and perioperative anesthesia management*. NY: Marcel Dekker, Inc, 2007;**1**:101–13.

Original research

26. **Wright MC**, Endsley MR. Building shared situation awareness in healthcare settings. In: Nemeth CP, ed. *Improving healthcare team communication: building on lessons from aviation and aerospace*. Aldershot: Ashgate Publishing Limited, 2008:97–114.
27. **Salas E**, Sims DE, Burke CS. Is there a « Big Five » in teamwork? *Small Group Res* 2005;**36**:555–99.
28. **Yule S**, Flin R, Maran N, *et al*. Surgeons' non-technical skills in the operating room: Reliability testing of the NOTSS behavior rating system. *World J Surg* 2008;**32**:548–56.
29. **Fletcher G**, Flin R, McGeorge R, *et al*. *Evaluation of the Prototype Anaesthetists Non-Technical Skills (ANTS) Behavioural Marker System*. Experimental Report (WP7). Scotland: University of Aberdeen, Department of Psychology, Industrial Psychology Group, 2002.
30. **Mishra A**, Catchpole K, McCulloch P. The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Qual Saf Health Care* 2009;**18**:104–8.
31. **Yule S**, Flin R, Maran N, *et al*. Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery* 2006;**139**:140–9.
32. **Yule S**, Flin R, Maran N, *et al*. Development of a rating system for surgeons' non-technical skills. *Med Educ* 2006;**40**:1098–104.
33. **Undre S**, Healey A, Darzi A, *et al*. Observational assessment of team work: a feasibility study. *World J Surg* 2006;**30**:1774–83.
34. **Frankel A**, Gardner R, Maynard L, *et al*. Using the communication and teamwork skills (CATS) assessment to measure health care team performance. *Jt Comm J Qual Patient Saf* 2007;**33**:549–58.
35. **Schraagen JMC**, Van de Ven J, Barach P, *et al*. Teams and cardiac surgery. In: Wong BLW, Stanton NA, eds. *Proceedings of the 9th Bi-Annual International Conference on Naturalistic Decision Making*. London, UK: The British Computer Society, 2009:152–9.
36. **Carthey J**. The role of structured observational research in health care. *Qual Saf Health Care* 2003;**12**(Suppl II):ii13–16.
37. **Baker DP**, Mulqueen C, Dismukes RK. Training raters to assess resource management skills. In: Salas E, Bowers CA, Edens E, eds. *Improving Teamwork in Organizations: Applications of Resource Management Training*. Mahwah, NJ: Lawrence Erlbaum Associates, 2001:131–45.
38. **Robson C**. *Real World Research*. 2nd edn. Oxford: Blackwell, 2002.