

The Editor-in-Chief welcomes topical correspondence from readers relating to articles published in BJS Open. Letters should be submitted via ScholarOne Manuscripts and, if accepted, will be published online.

Systematic review of methods to quantify teamwork in the operating theatre

DOI: 10.1002/bjs5.92

We welcome the attempt of Li and colleagues¹ to conduct a comparative analysis of the validity and reliability of teamwork tools for use in the operating theatre. This systematic review builds on and extends previous attempts to assess the psychometric quality of non-technical and teamwork assessment tools developed for use in the operating theatre; by bringing the tools together in a comparative commentary, this could provide a useful resource for those working in the area^{2,3}.

We read the systematic review with interest; we find that it adds to our understanding of the evidence base regarding teamwork in the high-risk perioperative environment. There are, however, some rather surprising omissions from this review, which we believe should be highlighted.

First, it is worth mentioning that the Non-Technical Skills (NOTECHS) instrument as revised from aviation was first published, with reliability evidence, by Sevdalis and co-workers⁴, but this reference has unfortunately been omitted from the review. Furthermore, a number of well known tools that quantify teamwork skills in the operating theatre were not included in the review. As stated, the aim of the review was to include teamwork assessment tools measuring teamwork of operating team members (not just surgeons). As such, the following behavioural rating systems that cover team skills are absent from the review: Anaesthetists' Non-Technical Skills (ANTS)⁵, Anaesthetic Non-Technical Skills for Anaesthetic Practitioners (ANTS-AP)⁶ and Scrub Practitioners' Non-Technical Skills (SPLINTS)^{7,8}. ANTS, developed in 2003, captures the non-technical skills

of anaesthetists, including task management, team working, situation awareness and decision-making. Evaluation of the ANTS system has provided data relating to both reliability and validity⁵. ANTS-AP, developed in 2015, captures the non-technical skills of anaesthetic assistants, including situation awareness, teamwork and communication, and task management. Evaluation of the ANTS-AP system has provided data relating to internal consistency, test–retest reliability, inter-rater reliability and accuracy⁶. SPLINTS, developed in 2013, captures the non-technical skills of scrub nurses, including situational awareness, communication and teamwork, and task management. Evaluation of the SPLINTS system has provided data relating to the reliability of SPLINTS⁸.

Second, we found that the evidence base for Observational Teamwork Assessment for Surgery (OTAS), one of the best-evidenced team assessment tools to date, was covered rather patchily in the review. A number of studies relating to the psychometric testing (reliability and validity evidence) of OTAS were not included^{9,10}. Furthermore, the culturally adapted Observational Teamwork Assessment for Surgery – Spanish (OTAS-S) was not included in the review, with its associated reliability and validity evidence¹¹.

Third, and perhaps most importantly, we are rather concerned with some of the commentary and interpretations of the reliability and validity metrics that the review reports on. Regarding reliability, the review offers no attempt to explain variation of inter-rater reliability in relation to the level of training of the raters who use the tools to carry out teamwork assessments. Assessment of teamwork is a skill within itself. As such, if observational behavioural rating systems are used in 'untrained hands', intraclass correlation coefficients (ICCs) are inevitably going to be low, whereas in 'trained hands' higher levels of agreement are to be expected. Thus, low to moderate ICCs between novices and experts at the start of training are expected and their improvement across the course of training reflects the importance of the training process,

not necessarily poor reliability of the tool itself. ICCs between trained assessors are the important metric when it comes to assessing the psychometric properties of a tool. The importance of non-technical and teamwork assessment training is shared among non-technical and teamwork skills tool developers, so much so that we have shown in an international Delphi study¹² that it ought to be taken into account before tool implementation. In addition, with regard to reliability, it is unclear why the authors state that 'scores should not be affected by testing at different sites' in reference to test–retest reliability. Different teams within different hospitals might reasonably be expected to display teamwork behaviours that differ in quality and receive different teamwork scores.

Regarding validity, we were surprised by the interpretation of a number of research findings in relation to the OTAS instrument – but also more broadly. Although the authors state correctly that previous research found that 'a proportion of OTAS components (behaviours or tasks) were consistently not witnessed in practice', they then go on to suggest that 'this may be explained by suboptimal team performance, but also casts doubt on its content validity'. We strongly disagree with the latter interpretation on two grounds. First, it is well recognized that team performance in the operating theatre is far from optimal and, considering that OTAS contains 'exemplar behaviours' (behaviours that indicate superior team performance), it is not surprising that operating teams do not display/engage in these superior teamwork behaviours during every operation. Second, by its very nature, observational assessment of perioperative teamwork will always be subject to the methodological limitation of how to interpret absence of a behaviour. Behavioural assessment tools, such as those the review covers, are not checklists, nor should they be (otherwise they would be remarkably unwieldy to use). Existing normative studies that offer consensus regarding the importance of behaviours covered by these instruments suggest that lack

of a behaviour is far from an indication of lack of validity; an example of this is the expert validation of OTAS exemplar behaviours¹³ and indeed the process of derivation of other instruments, such as ANTS and Non-Technical Skills for Surgeons (NOTSS).

A further point regarding validity is that, although teamwork requirements may be broadly similar across surgical specialties, we ought to remain aware that the needs of some of our colleagues in the operating theatre require further refinement of the available tools. There are surgical specialties in which the formation of the operating team includes team members that OTAS and NOTECHS simply do not capture. For example, we have recently developed a version of OTAS for use in endovascular surgery, Endo-OTAS¹⁴, as OTAS in its original form did not capture the team practices in this surgical specialty.

Overall, we found it interesting that limited forms of validity were included in the analysis. Considering that the results and discussion sections place a large focus on comparing OTAS and NOTECHS, we were surprised that the authors failed to present the data on convergent validity that exist between OTAS and NOTECHS, and are presented in one of the included articles: 'the overall agreement between OTAS and NOTECHS was excellent ($r = 0.886$, $n = 5$, $P = 0.046$)¹⁵.

After over 15 years of research on team and non-technical skills performance in the perioperative setting, reviews can help us take stock of the evidence base and direct future research, and also training and improvement efforts. However, the literature in this field is now quite expansive and review work ever more complex. Reviews with more narrow inclusion criteria and careful definitions of the subject matter being synthesized offer a useful and practical way forward.

Acknowledgements

S.R. is funded by a National Institute for Health Research (NIHR) Knowledge Mobilisation Fellowship.

The research of L.H. and N.S. is supported by the NIHR Collaboration for Leadership in Applied Health Research and Care South London at King's College Hospital NHS Foundation Trust. L.H. and N.S. are also funded by King's Improvement Science, which is part of the NIHR CLAHRC South London and comprises a specialist team of improvement scientists and senior researchers based at King's College London. Its work is funded by King's Health Partners (Guy's and St Thomas' NHS Foundation Trust, King's College Hospital NHS Foundation Trust, King's College London and South London and Maudsley NHS Foundation Trust), Guy's and St Thomas' Charity, the Maudsley Charity and the Health Foundation. N.S.'s research is further supported by the Economic and Social Research Council grant ES/P008313/1. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Disclosure: L.H., S.R. and N.S. have worked extensively on OTAS. N.S. is the director of London Safety and Training Solutions Ltd, which undertakes patient safety and quality improvement advisory and training services for healthcare organizations internationally.

L. Hull, S. Russ and N. Sevdalis
Centre for Implementation Science, Health
Service and Population Research
Department, Institute of Psychiatry,
Psychology and Neuroscience, King's College
London, London SE5 8AF, UK
e-mail: louise.hull@kcl.ac.uk

- Li N, Marshall D, Sykes M, McCulloch P, Shalhoub J, Maruthappu M. Systematic review of methods for quantifying teamwork in the operating theatre. *BJS Open* 2018; <http://doi.org/10.1002/bjs5.40>.
- Hull L, Arora S, Aggarwal R, Darzi A, Vincent C, Sevdalis N. The impact of nontechnical skills on technical performance in surgery: a systematic review. *J Am Coll Surg* 2012; **214**: 214–230.
- Sharma B, Mishra A, Aggarwal R, Grantcharov TP. Non-technical skills assessment in surgery. *Surg Oncol* 2011; **20**: 169–177.
- Sevdalis N, Davis R, Koutantji M, Undre S, Darzi A, Vincent CA. Reliability of a revised NOTECHS scale for use in surgical teams. *Am J Surg* 2008; **196**: 184–190.
- Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *Br J Anaesth* 2003; **90**: 580–588.
- Rutherford JS, Flin R, Irwin A, McFadyen AK. Evaluation of the prototype Anaesthetic Non-technical Skills for Anaesthetic Practitioners (ANTS-AP) system: a behavioural rating system to assess the non-technical skills used by staff assisting the anaesthetist. *Anaesthesia* 2015; **70**: 907–914.
- Mitchell L, Flin R, Yule S, Mitchell J, Coutts K, Youngson G. Development of a behavioural marker system for scrub practitioners' non-technical skills (SPLINTS system). *J Eval Clin Pract* 2013; **19**: 317–323.
- Mitchell L, Flin R, Yule S, Mitchell J, Coutts K, Youngson G. Evaluation of the scrub practitioners' list of intraoperative non-technical skills system. *Int J Nurs Stud* 2012; **49**: 201–211.
- Russ S, Hull L, Rout S, Vincent C, Darzi A, Sevdalis N. Observational teamwork assessment for surgery: feasibility of clinical and nonclinical assessor calibration with short-term training. *Ann Surg* 2012; **255**: 804–809.
- Sevdalis N, Lyons M, Healey AN, Undre S, Darzi A, Vincent CA. Observational teamwork assessment for surgery: construct validation with expert versus novice raters. *Ann Surg* 2009; **249**: 1047–1051.

- 11 Amaya Arias AC, Barajas R, Eslava-Schmalbach JH, Wheelock A, Gaitán Duarte H, Hull L *et al.* Translation, cultural adaptation and content re-validation of the observational teamwork assessment for surgery tool. *Int J Surg* 2014; **12**: 1390–1402.
- 12 Hull L, Arora S, Symons NR, Jalil R, Darzi A, Vincent C *et al.*; Delphi Expert Consensus Panel. Training faculty in nontechnical skill assessment: national guidelines on program requirements. *Ann Surg* 2013; **258**: 370–375.
- 13 Hull L, Arora S, Kassab E, Kneebone R, Sevdalis N. Observational teamwork assessment for surgery: content validation and tool refinement. *J Am Coll Surg* 2011; **212**: 234–243.e1–5.
- 14 Hull L, Bicknell C, Patel K, Vyas R, Van Herzeele I, Sevdalis N *et al.* Content validation and evaluation of an endovascular teamwork assessment tool. *Eur J Vasc Endovasc Surg* 2016; **52**: 11–20.
- 15 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**: 104–108.