

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

|                            |  |
|----------------------------|--|
| <b>TITLE (PROVISIONAL)</b> | Correlates of non-technical skills in surgery: A prospective study                                     |
| <b>AUTHORS</b>             | Gillespie, Brigid; Harbeck, Emma; Kang, Evelyn; Steel, Catherine; Fairweather, Nicole; Chaboyer, Wendy |

### VERSION 1 - REVIEW

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Akira Tsuburaya, MD PhD<br>Research Professor, Fukushima Medical University<br>Chief of Surgery, JCHO Nihonmatsu Hospital<br>1-553, Narita, Nihonmatsu, 9648501 Japan |
| <b>REVIEW RETURNED</b> | 16-Oct-2016   |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | <p>Major comments:<br/>This study uses NOTECHS for NOTSS evaluation and also as primary endpoints. Operative time, team membership, miscommunications, and interruptions are evaluated by NOTECHS. Since miscommunications and interruptions are directly correlated with NOTECHS subscales, partially equals, most of the results were a matter of course.<br/>For NOTSS evaluation, definitive endpoints are operative morbidity and mortality, and miscommunications and interruptions could be surrogate endpoints. I recommend adding a reverse analysis; miscommunications and interruptions as dependent variables, to identify improvable skills for each mistake.<br/>In multivariate analyses, operation time was not significant but the highest number of miscommunications was observed in cardiac surgery. As in discussion, the degree of difficulty, technical skills, stress, complexity, and patients' instability may be highest in cardiac surgery; the results may be affected by other factors.</p> <p>Minor comments:<br/>Page 11, L3. 19.1+/-3.5 may be wrong; change to 18.4+/-2.6<br/>Page 12, Table 1. Subscale 5 may be wrong; change to Subscale E</p> |
|-------------------------|---|

|                        |  |
|------------------------|--|
| <b>REVIEWER</b>        | Lene Spanager<br>Department of Surgery<br>Nordsjaellands Hospital<br>Dyrehavevej 29<br>3400 Hilleroed<br>Denmark |
| <b>REVIEW RETURNED</b> | 18-Oct-2016  |

|                         |  |
|-------------------------|--|
| <b>GENERAL COMMENTS</b> | I think this is a well-written manuscript on an interesting and clearly important topic. I just have a few comments:<br>1. Throughout the paper you use the term NOTSS. It refers to a |
|-------------------------|--|

|  |   |
|--|---|
|  | <p>published, well-defined taxonomy and rating scale for the observation of (individual) surgeons non-technical skills. You use the term as an abbreviation for the teams' non-technical skills. I would suggest that you use the abbreviation NTS (non-technical skills) instead, to avoid confusion.</p> <p>2. The introduction provides a good background and rationale for the study. However, it is slightly long and contains merely "classic" references on the topic. Perhaps you could find one or two newly published papers to add or replace some of the others?</p> <p>3. In the methods section you describe that interrater checks were made. I am curious to know if any rater training or calibration took place before the observations? And perhaps you could describe the raters' familiarity with the NOTECHS system?</p> <p>4. You recorded the miscommunications according to a framework published by Lingard et al. I think this adds validity to the results and would simply suggest that you refer to the framework in the methods section.</p> <p>5. I acknowledge the work put in the many observations - it is a clear strength of the study. You have good sampling across surgical specialties and you describe having data on 481 individual participants. However, it is difficult for the reader to judge whether the differences found between surgical teams can be attributed to different specialties or different surgeons, as the number of consultant surgeons and anaesthesiologists are not provided. Perhaps it could even reflect differences across the two hospitals, as the selection of specialties differed in the two hospitals. It is not a major thing, but I would suggest to discuss this in the paper.</p> <p>6. You found that miscommunications and interruptions affected the teams' non-technical performance (measured as NOTECHS scores), but operative time and familiarity did not. I would like to hear how you would explain this. And I cannot help to think that the reverse hypothesis could also be true: that miscommunications and interruptions (and perhaps operative time and familiarity) could impact the teams' non-technical skills. In fact, with this study design you cannot make sound conclusions on the causality of factors. I would suggest that you discuss this in the paper</p> |
|--|---|

### VERSION 1 – AUTHOR RESPONSE

| Reviewer comments  | Author responses  |
|--|---|
| <p>For NOTSS evaluation, definitive endpoints are operative morbidity and mortality, and miscommunications and interruptions could be surrogate endpoints.</p> <p>I recommend adding a reverse analysis; miscommunications and interruptions as dependent variables, to identify improvable skills for each mistake.</p> | <p>We carefully considered rerunning the analysis to make the predictor variables dependent variables. We recognise that the design of the study does not allow the temporal order between predictors and outcomes to be determined; therefore, causal relationships cannot be established. Consequently, there may be some competing explanations for these results. This limitation has been acknowledged in the Discussion section, pp. 12-13, lines 275-80.</p> <p>Our hypothesised predictors of NTS were identified in the growing body of literature that suggests factors such as interruptions, miscommunications and team familiarity potentially erode surgical teams' NTS. For example:</p> |

|   |   |
|---|---|
|   | <p>Kurmann A, Keller S, Tschan-Semmer F, et al. Impact of Team Familiarity in the Operating Room on Surgical Complications. <i>World J Surg</i> 2014;<b>38</b>:3047–52 doi: DOI 10.1007/s00268-014-2680-2[published Online First: Epub Date]].</p> <p>Gillespie B, Chaboyer W, Fairweather N. Interruptions and miscommunications in surgery: An observational study. <i>AORN Journal</i> 2012;<b>95</b>(5):576-90</p> <p>Campbell G, Arfanis K, Smith A. Distraction and interruption in anaesthetic practice. <i>British Journal of Anaesthesia</i> 2012;<b>109</b>:707-15 doi: 10.1093/bja/aes219[published Online First: Epub Date]].</p> <p>Kang E, Massey D, Gillespie B. Factors that influence the non-technical skills performance of scrub nurses: A prospective study. <i>J. Adv. Nurs.</i> 2016;<b>71</b>:2846–57 doi: 10.1111/jan.12743[published Online First: Epub Date]].</p> |
| <p>In multivariate analyses, operation time was not significant but the highest number of miscommunications was observed in cardiac surgery. As in discussion, the degree of difficulty, technical skills, stress, complexity, and patients' instability may be highest in cardiac surgery; the results may be affected by other factors.</p>                                 | <p>Thank you, we have elaborated on these points in the Discussion section (pp. 11, lines 247-251).</p>   |
| <p>Page 11, L3. 19.1+/-3.5 may be wrong; change to 18.4+/-2.6</p>   | <p>The symbol <math>\pm</math> has been used to indicate standard deviation. We will be guided by the Editor in relation to whether there is a preference in the way this is written for BMJ Open.</p>  |
| <p>Page 12, Table 1. Subscale 5 may be wrong; change to Subscale E</p>  | <p>Revised accordingly.</p>   |
| <p>Throughout the paper you use the term NOTSS. It refers to a published, well-defined taxonomy and rating scale for the observation of (individual) surgeons non-technical skills.</p> <p>You use the term as an abbreviation for the teams' non-technical skills. I would suggest that you use the abbreviation NTS (non-technical skills) instead, to avoid confusion.</p> | <p>The abbreviation NOTSS has been changed to NTS throughout the paper.</p>   |
| <p>The introduction provides a good background and rationale for the study.</p>   | <p>I have added the following more up-to-date references:</p>   |

|   |   |
|---|---|
| <p>However, it is slightly long and contains merely "classic" references on the topic. Perhaps you could find one or two newly published papers to add or replace some of the others?</p>   | <p>Raman J, Leveson N, Samost A, et al. When a checklist is not enough: How to improve them and what else is needed. <i>The Journal of thoracic and cardiovascular surgery</i> 2016;<b>152</b>(2):585-92.</p>   |
| <p>In the methods section you describe that interrater checks were made. I am curious to know if any rater training or calibration took place before the observations? And perhaps you could describe the raters' familiarity with the NOTECHS system?</p>  | <p>I have added further description in the Methods section, pp.4, lines 84-89.</p>  |
| <p>You recorded the miscommunications according to a framework published by Lingard et al. I think this adds validity to the results and would simply suggest that you refer to the framework in the methods section.</p>   | <p>I have included several citations of Lingard et al, and Healey et al's frameworks, pp. 6, lines 134-140 in the Methods section.</p>  |
| <p>I acknowledge the work put in the many observations - it is a clear strength of the study. You have good sampling across surgical specialties and you describe having data on 481 individual participants. However, it is difficult for the reader to judge whether the differences found between surgical teams can be attributed to different specialties or different surgeons, as the number of consultant surgeons and anaesthesiologists are not provided. Perhaps it could even reflect differences across the two hospitals, as the selection of specialties differed in the two hospitals. It is not a major thing, but I would suggest to discuss this in the paper.</p> | <p>I agree that it is difficult to judge whether differences b/ teams can be attributed to different surgeons/other staff members or different specialties. Given the differences in the types of specialties, we can only speculate why some of these differences might exist. We have acknowledged the limitations of the design on pp. 12-14, lines 275-80.</p> <p>I have added some discussion around these points on:</p> <ul style="list-style-type: none"> <li>• pp. 11, lines 247-58</li> <li>• pp. 12, lines 259-71</li> </ul> |
| <p>You found that miscommunications and interruptions affected the teams' non-technical performance (measured as NOTECHS scores), but operative time and familiarity did not.</p> <p>I would like to hear how you would explain this. And I cannot help to think that the reverse hypothesis could also be true: that miscommunications and interruptions (and perhaps operative time and familiarity) could impact the teams' non-technical skills. In fact, with this study design you cannot make sound conclusions on the causality of factors. I would suggest that you discuss this in the paper</p>  | <p>We recognise that the design of the study does not allow the temporal order between predictors and outcomes to be determined; therefore, causal relationships cannot be established. Consequently, there may be some competing explanations for these results. This limitation has been frankly acknowledged in the Discussion section, pp. 12-13, lines 275-80.</p>   |

## VERSION 2 – REVIEW

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Akira Tsuburaya, MD PhD<br>Gastrointestinal Surgery, Fukushima Medical University<br>1, Hikarigaoka, Fukushima City, 9601295, Japan |
| <b>REVIEW RETURNED</b> | 04-Nov-2016   |

|                         |  |
|-------------------------|--|
| <b>GENERAL COMMENTS</b> | The response for the reviewers and the revision are appropriate, and well done.<br>I think it can be published without further revision. |
|-------------------------|--|

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Lene Spanager<br>Department of Surgery, Nordsjaellands Hospital, Hilleroed, Denmark |
| <b>REVIEW RETURNED</b> | 10-Nov-2016   |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | I have now had the privilege to read the manuscript twice and I think it has improved considerably with the recent changes. In fact, I only have one comment: reference #21 and #22 appear to be identical. |
|-------------------------|---|

## VERSION 2 – AUTHOR RESPONSE

- Duplicate reference (#21 and #22) has been removed.