Author's response to reviews

Title: Surgical Frailty Assessment: A Missed Opportunity

Authors:

Gilgamesh Eamer (eamer@ualberta.ca)
Jennifer Gibson (jennagibson@gmail.com)
Chelsia Gillis (chelsia.gillis@ucalgary.ca)
Amy Hsu (ahsu@toh.ca)
Marian Krawczyk (mrnkrawczyk@gmail.com)
Emily MacDonald (emily.macdonald@unb.ca)
Reid Whitlock (rwhitlock@sogh.md.ca)
Rachel Khadaroo (khadaroo@ualberta.ca)

Version: 2 Date: 19 Jun 2017

Author’s response to reviews:

The following responses are also available in a formatted form as an attached document.

Dear Dr. Faraoni

We would like to thank the editors and reviewers for their valuable feedback and constructive criticism. We have prepared responses to the comments from each reviewer below and revised our manuscript to address their concerns. We hope these improvements are well received and look forward to your response.

Gil Eamer, MD

Reviewer 1:

1. I can't find "HCP" spelled out.

Response: We have added a definition to the main body of the text (page 3, paragraph 3). HCP refers to healthcare professionals and is also defined in the abbreviations section.
2. The sample size - both of surgeons and allied health staff is small, and a single centre. However, it may actually reflect more positively than a wider survey given this centre is undertaking a study on frailty.

Response: Thank you for the critique recognizing that the study findings might be more positive considering that the single-site study sample was apart of a broader frailty initiative (i.e. the larger, on-going Elder-friendly Approaches to the Surgical Environment (EASE) Project). We hope that the findings here will inform and add to the interpretation of the results from that analysis.

We have used this critique to edit the discussion section significantly to better articulate how the findings were appreciated in relation to the larger study. We have added to our limitations section discussing the single centre nature of our survey and its effect on generalizability. We expect that our findings may actually reflect more positively than these other centers.

3. This work reflects much of what is already know about frailty - i.e. its importance, but its lack of incorporation into routine operative planning. Thus, I would be more interested in any results that looked at 1. What were the "unique" challenges of the surgical setting. 2. How identifying a frail patient changed pathways/planning for each of the health professional groups. 3. And trying to re-frame the article around these. These barriers are the most interesting component of the study.

Response: Thank you for this suggestion. While we are aware that this “knowledge-do” gap has been previously documented, we believe our study results offer new insights – that even when frailty assessments are incorporated into routine care planning, there remains significant barriers to consider. Our finding also further support what is already known – i.e., while there was a positive correlation between the belief that “frailty assessment should be done for all surgical patients” (as a proxy measure for the importance of identifying frail patients) and the likelihood of incorporating frailty in their own practice/care planning for patients. This finding was consistent for all health professional groups. We have added this perspective to our discussion.

The authors agree that discussion of the study findings with particular explanation about barriers are of interest to readers. We have expanded on our qualitative analysis to provide quotations from respondents that provide examples of the barriers to care of the frail that we identified that were specific to the surgical setting in the results subsection titled ‘Barriers to frailty assessment’ (Pages 8-9, all paragraphs). Given that nature of the research (survey) design, we are unable to comment on how respondents changed their care practices.

Reviewer 2:

In this single-centre, observational study, the authors surveyed health care professionals (HCP) caring for patients enrolled in the Elder-friendly Approaches to the Surgical Environment (EASE) study in an effort to assess perceptions and attitudes towards perioperative frailty assessment. Responses to an author-developed survey were obtained from 49 of 117 (42%) HCP from three main subgroups: nurses, surgeons, and other allied health professionals. The authors
report significant intergroup differences in responses. Lack of knowledge about frailty issues was a prominent barrier to the use of frailty assessment in practice across all subgroups of HCPs.

For the most part, the manuscript is well-written. I do have a couple of comments about the survey development and analysis:

Response: Thank you

1) During survey development, there appears to be an attempt to ascertain content validity. How did the authors assess reliability? This should be described in the methodology section.

Response (1): During survey development, we assessed content validity by seeking feedback from other experts in the field and by using a cognitive interviewing approach. The initial draft of the survey was pilot tested with 2 physicians, 2 RNs and 2 allied health professionals to determine the survey’s construct validity, interpretability, redundancy and ease of administration. Reliability was assessed using Cronbach’s alpha. We have now added this to the manuscript in the methods section (Page 5, paragraph 1), and results section (Page 6, paragraph 1).

Reviewer follow-up: Please provide a more detailed description about the assessment of reliability. Stating that Cronbach’s alpha was used is inadequate and doesn't provide insight into the actual methodologic approach. Was a test-retest methodology used? If so, please describe. Did the authors assess inter-rater reliability? If so, was inter-rater reliability assessed by HCP subtype? A more detailed description of the approach is warranted.

Response (2): As described in the ‘Material and Methods’ section (Page 4, paragraph 3), we followed the recommendations of Willis et al. (2013) – “What Do Our Respondents Think We’re Asking? Using Cognitive Interviewing to Improve Medical Education Surveys” – when we had designed and evaluated our survey using the cognitive interviewing technique to enhance validity. When evaluating reliability, we identified themes within our survey results and used Cronbach’s alpha to test reliability for these two main themes along with the overall study results. We recognize that adding questions can artificially increase the alpha without increased reliability for our overall study alpha calculation. Consequently, we have reported alpha for our two main themes along with the overall alpha (page 6, paragraph 3). Our first theme (frailty assessment) remains acceptable while our second, and less important theme, (knowledge about the frailty scale and the EASE study) has marginal reliability. This has now been reported in the results. We have also described the predetermined cut-offs for alpha that we used (page 5, paragraph 1).

We did not conduct a test-retest analysis due to the busy schedule of our survey respondents and the physician distributing the survey. Inter-rater reliability in survey results are typically used when surveys are performed in person, the team member conducting the survey is required to rate the survey respondent on one or more scales and there is more than one team member conducting the survey. Inter-rater reliability is not relevant to our survey results as there are no ratings that were recorded by the survey team and the survey was distributed and collected by one person.
2) It would be interesting to see all of the final 26 survey items included. These are supposed to be present in Figure 1 (page 5) but I am unable to see a Figure 1 with the submission.

Response (1): It is now included in Appendix 1 as a separate attachment

Reviewer follow-up: I am unfortunately still unable to open or see Appendix 1.

Response (2): We have attempted to re-attach the appendix (Survey instrument). If you remain unable to see appendix 1, please contact the editors for assistance.

3) Mean response scores as a function of HCP subgroup to a subset of survey items are presented in Table 2. The authors further describe results of t-tests reflecting pairwise comparisons of two of the three subgroups (i.e. page 6 - nurses vs. surgeons, allied health vs. surgeons). Was there an attempt to control for multiple comparisons during the analysis (eg. ANOVA, followed by post-hoc tests?) If so, the details of this should be described in the analysis section.

Response (1): During our analysis, we had considered performing multiple comparison using ANOVA. However, given the small sample size, there were concerns about non-normality and inequality of variances. In fact, most responses did appear to have a non-normal distribution (overall and by HCP subgroups). As a result, we suspected that calculated F statistic may be dominated by the sample variances for the larger samples and the test will not correctly identify significant differences in the means. Therefore, we chose to conduct t-tests and report these results in the paper. We did compare nursing and allied health using t-tests as well and did not find a significant difference. We have added a sentence (Page 7, paragraph 1) to reflect this.

Reviewer follow-up: The authors state that concerns about normality, data distribution, and sample size precluded an inferential approach with ANOVA and they proceeded to conduct pairwise t-tests instead. Like ANOVA, inferential testing with a two-sample t-test necessarily assumes that the data is normally distributed and that the variances of the two populations are equal (please see https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Two-Sample_T-Test.pdf). As such, given the aforementioned concerns about the data, this is an inappropriate test to use in this instance.

Response (2): Thank you for the reference. We agree that both methods depend on the assumption that the data is approximately normal. The main reason why we have decided to present the t-test instead of ANOVA results, however, is because of our small sample size. It is our understanding that the F-test, which ANOVA is based on, is more sensitive to nonnormality and less likely to detect true distributional differences when the sample size is small (https://www.rand.org/content/dam/rand/pubs/research_memoranda/2008/RM5072.pdf), while the t-test has been shown to be valid for samples that are as small as 2 observations (http://pareonline.net/getvn.asp?v=18&n=10).

Despite this, we recognize the relevance of multiple comparisons and the value of ANOVA with post-hoc tests. We have also clarified in our methods section that we conducted pairwise t-tests for our main comparisons and ANOVA for our sensitivity analysis (Analysis subheading, page 5, paragraph 3 and Appendix 2) and we describe the findings in our results section (page 10).
4) Interpretation of Table 2 is confusing. For instance, 16 nurses were surveyed but it appears that only 9/13 answered the first item? How should the '9' and '13' be interpreted? Some explanatory table footnotes would be helpful.

Response (1): Not all respondents answered all questions. Each ratio represents all people who strongly agreed or agreed (numerator) divided by the number of responses to that question (denominator). The mean score (1=“Strongly disagree”, 3=“Neither agree nor disagree”, and 5=“Strongly agree”) is then reported below. A new footnote has now been added to all three tables to explain this.

Reviewer follow-up: Thank you for the clarification and the addition of the table footnotes. What were some of the reasons for substantial variability in non-responses to the various survey items?

Response (2): Thank you for your important question. In our survey we did not collect reasons for non-responses. However, the largest number of non-responses was in our section in the section exploring HCPs’ perceived usefulness of the CFS score across disciplines/professions (Table 3).

Despite our efforts to ensure that the survey items are easily interpretable and have minimal redundancy, we suspect that some of the items may be difficult to provide a response due to the lack of knowledge about or unfamiliarity with frailty assessment tools. For example, it may be difficult for the study participants to say whether they have “…always use a frailty assessment tool to assess patients for frailty” when they don’t know what qualified as a frailty assessment tool. Despite being part of the EASE study, many were unaware of the CFS. As noted in our manuscript, many health professions included in our study continue to use clinical judgement to determine a patient’s frailty. We believe this may have contributed to the larger proportion of non-responses to items that attempted to measure their use of frailty assessment tools.

Of the 23 HCPs who did not respond to the 2 questions: “The CFS score is useful to the overall perioperative care that is provided in the hospital” and “The CFS Score is useful to the perioperative care that I provide in the hospital” 22 (96%) of them indicated they were not aware that the Clinical Frailty Scale (CFS) is a frailty assessment tool. Of the 18 HCPs who did not provide a response to “I would like to use or continue using the CFS score in my care of older adults” 16 (89%) of them indicated they were not aware that the CFS is a frailty assessment tool.

Furthermore, when exploring HCPs’ perceived need for additional frailty assessment and training score across disciplines/professions (Table 4). Eleven fewer respondents answered whether they would benefit from further training on the CFS compared to whether they would benefit from further training on assessing frailty in general. Of the 11, 10 indicated they were not aware that the CFS is a frailty assessment tool.

Likert scale questions related to the CFS were the most likely to have a non-response and the non-responders were highly likely to not be aware of the CFS. There were comparatively few non-responses related to Perceived importance of frailty assessment across disciplines/professions (Table 2). Therefore, we have reason to believe that is was lack of awareness of the CFS as a frailty assessment tool that was a primary driver of substantial
variability in non-responses to the various survey items. We have addressed this point in the limitations (page 13, paragraph 1).

5) The authors report selected Spearman's rank-order correlations between various survey items. Was there ever consideration given to examining these correlations as a function of HCP subgroup which can be accomplished with a linear regression framework?

Response (1): We have considered taking a regression approach to examine the relationship between various provider characteristics (i.e., age, sex, HCP subgroup, years of experience in profession, and years of experiences in a surgical setting) and the observed differences in the responses (i.e., the perceived importance of frailty assessment and the perceived usefulness of the CFS score). However, the models demonstrated poor fit ($R^2$ value < .1) and the model F-tests suggested these characteristics were not significantly associated with the reported differences in the outcomes. However, we suspect that the lack of significant differences in our multivariate models was a result of the small sample size and insufficient power. Therefore, we chose to examine the data using bivariate correlation analysis.

Reviewer follow-up: Thank you for the response.

Response (2): We are pleased that we were able to answer your concern to your satisfaction.

6) On page 7, the authors state 'Four survey questions captured interdisciplinary HCPs perceptions of the usefulness of frailty assessment scores in clinical practices (Table 3). I only see 3 questions with their mean responses in Table 3. Should the fourth be one relating to awareness of the CSF?'

Response (1): Three questions were Likert scale questions (Table 3) and one was open-ended. This has been clarified in the text (‘Survey results’ subsection, page 7, paragraph 3, line 1). The open-ended responses are discussed in the next paragraph.

Reviewer follow-up: Thank you for the response.

Response (2): We are pleased that we were able to answer your concern to your satisfaction.

7) On page 6, what do the authors mean by 'we conducted tests of independence in the pooled surgeon subgroup'?

Response (1): We wanted to ensure that pooling resident physicians with experienced surgeons was not inappropriate (maybe additional experience led to differing practices with respect to frailty). We tested to ensure each subpopulation didn’t differ from each other in terms of response patterns. This has been clarified in the results section (‘overall’ subsection, page 6, paragraph 1).

Reviewer follow-up: What statistical test(s) were used to conduct these tests of independence?
Response (2): Thank you for the opportunity to clarify. We calculated the mean response scores, frequencies, and percent “agree” or “strongly agree” for each survey item for both staff surgeons and surgical residents. Comparison between these two subgroups was conducted using the t-test for independent samples. We have added this information to the Overall subheading in the Results section. (page 6, paragraph 1)