Author's response to reviews

Title: The impact of performance status and comorbidities on the short-term prognosis of very elderly patients admitted to the ICU

Authors:

Fernando G Zampieri (fgzampieri@gmail.com)
Fernando Colombari (fernando.colombari@haoc.com.br)

Version: 3 Date: 16 May 2014

Author's response to reviews: see over
We would like to thank the reviewers for their thoughtful comments on our manuscript. We provide below, in bold, a point-by-point answer to the reviewer’s queries. Changes in the manuscript have been highlighted in red. We would like to highlight that minor changes in the results occurred after the reviewer’s comments. These changes are all marked in red in the Tables and in the manuscript. We also improved the recursive partitioning model with slightly different results. Nevertheless, the core results and the meaning of our findings were left unchanged after this revision.

**Reviewer 1:**

Zampieri and Colombari have conducted a retrospective analysis of patients 80 years of age and older in their ICU to identify predictors present on admission to the ICU of hospital mortality. I find the study novel in its focus on the extreme elderly and its inclusion of prior markers of illness (performance status and comorbidities). Their finding that the premorbid condition of the extreme elderly is relevant to outcomes is interesting. My main concern regards the presentation of the results and the discussion which I feel, at times, lack focus and draw attention away from the results and how to place them in context of previously published studies.

Thank you for your peer review. We agree that this subject is interesting due to the growing number of very elderly patients admitted to the ICU. We hope that we
were able to address the reviewer’s concerns regarding the discussion and presentation of the results, as discussed below.

Major Comments

1. Focus of results section – While the main results pertain to the multivariable model and conditional inference tree, the first 3.5 out of 6 paragraphs in the results section describe the cohort. I find there to be too much detail about the cohort itself (especially the bit referring to table 1 in which patients are not stratified by mortality) and too little time is spent exploring the details of the main results which seem to begin in paragraph 4 of the results with “After multivariate analysis…”.

Thank you bringing this to our attention. We simplified the first paragraphs of the results section. Additional information was added to Table 3 that was now renamed as Table 1, as suggested. The previous Table 1 was changed to a supplementary table in the additional file. We further expanded the presentation of our findings in the Results section.

2. The discussion section does not fully help me to understand how to interpret their results as a package – The 2nd and 3rd paragraphs (which discuss the results of the multivariable model) are hard for me to follow. It would be helpful to have a more clear explanation here on how the major points of the model fit into existing literature (this is what paragraph 2 is aimed at doing, I think, but it is hard for me to get a good sense from). Paragraph 3 is a very detailed (too detailed, in my opinion) explanation of the statistical properties of the model they have created; I am not sure whether this is necessary in this depth. Finally, there is no explanation of how to interpret the model AND the inference tree together. I find the technical details explored in the discussion
less important and feel that it is missing a chance to tell the reader how to really understand the findings.

Thank you for bringing this to our attention. We agree with the reviewer that the second paragraph of the discussion was confusing and did not inform the reader of the relevance of our findings. We rewrote the second paragraph in order to better situate the reader and where our study fits. We also added a sentence at the beginning of the fifth paragraph that aims to synthesize the link between the traditional logistic regression and recursive partitioning and summarizes our findings. This resulted in a much improved version of the discussion and increased reading flow.

Regarding the statistical detail presented in Paragraph 3, we disagree with the reviewer that excessive technical information is presented. We believe technical aspects of the analysis that could influence the results should be presented, especially when a retrospective analysis is performed. The calibration issue is prominent, since it is frequently overlooked. There are a number of ways to assess calibration and model properties, each with its unique features. We believe that relying on Hosmer-Lemeshow test to assess calibration is inadequate. Since one of the most unique aspects of our analysis is the use of the calibration plot after bootstrap, we felt that its use should be discussed on the Discussion.

3. Are these results unique to older populations? – I do not know whether the results found herein are unique to older populations, but I would consider adding a section to the discussion (if this is known – ie, whether PS in particular is a predictor of outcome in other age groups) or, if it is not known in the literature, to the results as a separate, similar, analysis in younger patients. Currently the manuscript suggests PS and
comorbidities add important prognostication in extreme elderly, but I wonder whether this is true in younger patients as well.

This is a very interesting question! Performance status has been repeatedly shown to be associated with prognosis in critically ill oncologic patients, for example. We have also previously shown that PS is independently associated with higher ICU LOS (Zampieri FG, JCC 2014). Similarly, measurements of frailty have been associated with worst outcomes in critically ill patients (Bagshaw, CMAJ 2014). We are currently performing a study that aims to answer exactly the same question proposed by the reviewer: Is Charlson Comorbidity Index and PS important in younger patients? Nevertheless, we display below a preliminary analysis of our data for the reviewer. We performed an analysis on patients with less than 65 years. Data was retrieved from the same database we used for the current manuscript. Please note that this is an unpublished analysis and we are waiting for more patients to be included on the database because there are few young patients with PS of 2. This analysis used exactly the same statistical plan shown on our manuscript. For patients under 65 years (1790 patients), SAPS 3, Charlson Comorbidity Index and PS were independently associated with hospital mortality. OR are 1.12 [1.10-1.14] for SAPS 3, 1.29 [1.17-1.42] for Charlson and 2.84 [1.49-5.41] for a PS of 1 (need for assistance). PS of 2 was not associated with worst outcome, probably because there were too few patients with PS of 2 below 65 years (less than 2%). This analysis, therefore, is clearly underpowered.

Since this is an ongoing study, we did not mention this on our manuscript. Nevertheless, as suggested, we added to the discussion the currently published
results by Bagshaw, Soares among other authors that associated worst PS with worst prognosis.

Minor Comments

Abstract

• Results section: “…were associated with partially dependent patients…”: To which outcome does this refer? The second portion of the sentence says “while… was associated with increased hospital mortality”—is the first half referring to a different outcome?

Thank you very much for bringing this to our attention. The Results section in the abstract was really confusing and in disagreement with the study results. It was now rewritten.

Introduction

• Paragraph 2, line 2, “Despite… The scores from this system… Moreover…” – I find these 3 sentences difficult to understand. In the first it is said that the discrimination is good, but then the next 2 speak to problems with it. Perhaps this could be clarified.

We tried to clarify this paragraph. Now it clearly state that despite an overall good discrimination capability of the SAPS 3 score there are doubts regarding its calibration. We hope that this change will improve clearness of the manuscript.

Methods

• It would be helpful to have the specific independent variables considered for the multivariate model outlined somewhere in the methods section.

This information was added to the methods section.
• 2nd to last paragraph, “Cutoffs… were automatically selected.” – Does this mean that the algorithm selected those which were most useful, they were selected arbitrarily by the algorithm, or some other strategy?

In recursive partitioning using the *party* package of R project, the cutoffs are automatically chosen by the internal procedures of the package in order to provide the best discriminative capability of the model.

Results

• As mentioned in the major comments, I would consider shortening the first 3.5 paragraphs. I am not sure what table 1 adds to table 3 and I would consider making table 3 into a table 1.

This paragraph was rewritten and Table 3 and 1 were merged, as suggested. This greatly improved the manuscript and we would like to thank the reviewer for his comment.

• Paragraph 4, “After multivariate analysis…(Table 4); these variables were used for the final model.” – What is the final model—is it different from the multivariate model? I am not sure, but perhaps there is a multivariate model made of all significant independent variables as determined by the backwards stepwise process and then only those that are significant in this model are included in a final one? Please make this more clear in the methods and then here in the results as well.

This was indeed confusing on the manuscript. The final model incorporated all variables after stepwise regression and not only those statistically associated with outcome. The only variable that was present in the final model but that was not associated with outcome on logistic regression was sepsis, with a marginal p = 0.07.
Triggered by the reviewer comment, we excluded sepsis from final model and kept only variables associated with outcome. This slightly changed the values of the OR and calibration plot, but did not alter the essence of the manuscript. AUC was unaltered.

• Paragraph 5, line 7, “In essence… of approximately…”: should the “of “ be an “at” instead? The errors are at probabilities 0.3 and 0.5, they are not errors of this magnitude, I think.

The reviewer is correct. This was changed in the manuscript as suggested.

Discussion

• Paragraph 1, line 5, “…and the number of ICU admissions…”: this is a model of individual patient characteristics, I think. As such, I am not sure what this phrase means. Does it mean the number of sepsis admissions in the ICU to which the patient was admitted? Please clarify.

We clarified this on the manuscript. It meant admission due to sepsis. Thank you.

• Paragraph 2, lines 4-5, “the only predictor of mortality… were the only predictors of long-term mortality.”: Is the first reference to mortality meant to be short-term mortality? As it reads now, I am not sure what the difference is between it and long-term mortality.

The paragraph was rewritten as suggested previously. There is a wide variation in the definition of “long term” in the literature; each study defined long term its own way. Short term refers to hospital mortality on most studies.
No, because there was no overestimation in the referred situation, only a reduction in the underestimation after adding Charlson, PS and non-full code status to the model. This paragraph refers to the calibration plot. We tried to clarify this on the manuscript.

Tables/Figures

- Table 1: I do not know that this is needed; I would consider making the current Table 3 into a Table 1 with inclusion of all variables listed in Table 1. Is “non-full code status” just on admission to the ICU or at any time during the stay?

This was performed as requested.

- Table 4: It would be helpful to know what other variables were included in the final model? (or is that it? – if so, then those considered but not included.)

It is shown marked on the table which variables were included for logistic regression.

- Figure 3: It is hard to make out the different lines on the graphs; can only clearly make out 2 (I think ideal and bias-corrected).

Unfortunately, this is because the two lines are really close one to another. Due to software limitations (the calibration plot in R) we are unable to perform changes in the figure in order to clarify this.

We would like once again to thank the reviewer for his detailed peer review. We are sure that the manuscript was much improved after his comments.
Reviewer 2:

The manuscript: "The impact of performance status and comorbidities on the short-term prognosis of very elderly patients admitted to the ICU" is a retrospective cohort analysis of mortality in a very elderly population admitted to the ICU. The work focuses on determining the discrimination and calibration characteristics of the SAPS 3 model in this population as well as determining whether the addition of a Charlson Index and/or Performance Status would improve these characteristics. The article is, for the most part, well written and clear.

Examining prognostic models in the very old is of clinical relevance as traditional models are validated on populations with a more disparate age range and palliative care discussions are, in general, more common during the care of elderly patients. However, there are some important features of this manuscript that should be addressed.

Thank you for your detailed peer review. We provide an answer to your queries below.

Major:

1) While the addition of the Charlson Index and Performance Status may help to risk adjust for future research studies, I struggle to see how this work will realistically assist in clinical care. It is uncommon for clinicians to formally calculate SAPS 3 scores on patients prior to "goals of care" discussions and pragmatically speaking, it is unlikely that clinicians will also calculate a Charlson score as well. Does the modest improvement in the ROC AUC justify the increased work required to generate it? One way to potentially maximize the clinical relevance of this work would be to look at individual components of the SAPS 3 and Charlson score and identify which factors have the strongest association with outcomes in the population. This may result in a
much more "user friendly" scoring system for the clinician as he or she prepares to discuss goals.

The reviewer poses an interesting remark regarding our manuscript: Despite being able to show that PS and Charlson are associated with short-term prognosis, the improvement in accuracy is small, so is our model useful at the bedside, especially regarding end-of-life discussion? We would like to make some additional comments.

We agree that physicians are usually reluctant to calculate several index and scores at the bedside, especially because there is a profusion of index and scores. We are not suggesting that the physician should calculate Charlson, measure PS and use it to discuss goals of care. We aimed to describe the association between such variables and short-term prognosis, and suggested that their absence on a popular prognostic score may be the cause of their moderate accuracy and poor calibration on specific probability strata.

Therefore, the question our manuscript tried to answer was: Are burden of comorbidity and performance status associated with short-term outcome in critically ill elders? If so, could the addition of these values to SAPS3 reduce the reported calibration problems SAPS3 usually present and improve its accuracy? Our answer to both questions is yes, and our work should be viewed as a hypothesis generating study. Maybe future prognostic scores, adequately created using prospective collected data, should include at least simple background information.

Nevertheless, triggered by the reviewer’s interesting suggestion, we explored the comorbidities associated with poorer outcome and proposed a score,
named Elder Performance and Comorbidity Prognostic Score (EPCP Score). The score was constructed using logistic regression and included the same variables included on our final model, exchanging Charlson to the comorbidities associated with hospital mortality. This score had good accuracy (similar to SAPS 3 but lower than the final model) and good calibration (but with more systematic errors than our final model). This highlights that baseline features and reason for admission are extremely important determinants of hospital mortality, but a measurement of illness severity still play an important role. This score was added to the manuscript, with detailed information on the newly added Additional File.

2) Bias is a concern here. I am particularly worried about the association with "non-full code" status and mortality. It is very possible (if not likely) that many of those who have "non-full code" status were given this status only after it became apparent that they were not responding to therapy and had a small chance of survival. Thus, once it looked like they weren't going to survive, they were made DNR. As such, "non-full code" status almost becomes like a self-fulfilling prophecy for mortality. If there were a way to determine if a patient was "non-full code" at the time of admission, this would be, perhaps, a more relevant association. If not, this bias needs to be clearly outlined in the limitations section.

Thank you for your comment. The reviewer is correct regarding the eventual presence of bias in our analysis in the non-full code group. We added this to the limitation paragraph (last sentence). Please note that, according to Conditional Inference Tree, the impact of a non-full code status was only present on patients high SAPS 3 and was associated with an increase in mortality from 76 to 86%. Also, to highlight that the other associations regarding PS and Charlson are valid regardless of non-full code status, we performed a sensitivity analysis excluding all
patients that had a non-full code status. We added this analysis to the Additional File and we will discuss it below after your third remark.

3.) Similarly, there likely also exists bias in regards to someone's performance status. If it is known, up front, that a patient has a poor baseline performance status, clinicians may be biased in how they approach care and family meetings which could then lead to higher mortality rates among those with a Performance Status of 2. For example, it is noteworthy that only 15% of the population received mechanical ventilation while published estimates have shown much higher rates in ICUs (Wunsch et al. Crit Care Med 2013). Are some patients who would ordinarily receive mechanical ventilation receiving NIPPV instead (use rate = 17%) due to clinician bias? Alternatively, this low rate of mechanical ventilation use may be more reflective of institutional care practices, but, either way, affects the generalizability of the current results.

Our hospital is a tertiary hospital without a step-down unit or a high dependency unit. On a previous study by our group performed on the same unit, use of organ support was similar (Zampieri, Journal of Critical Care, 2014). Therefore, we believe that this low use of support was a characteristic of our unit and not a sign of life support limitation. Nevertheless, as mentioned above, we performed a secondary analysis excluding patients with a non-full code status. Results are shown on sTable 5 on Additional File and highlight that except for PS 1 (which had a p value of 0.052 and therefore almost reached significance) all other variables were kept in the model with similar OR and confidence intervals. Therefore, we do not believe that our findings can be solely explained by bias.

Minor:

1) I would try to confine conjecture to the discussion section alone (Figure 4 legend)
This was performed as suggested by the reviewer.

2) Introduction is a bit disjointed. Paragraph 2 talks about calibration concerns of SAPS3, then jumps to its discrimination then jumps back to calibration.

This paragraph was rewritten after Reviewer 1 and Reviewer 2 comments.

3.) The Results section of the Abstract needs to be clearer as it pertains to the OR for each score on the performance status.

The Result section of the abstract was rewritten.

4.) The Results text and Table 2 are not congruent. One says NIPPV = 17% and the other 15% etc.

Incongruences were solved in this new version of the manuscript. Thank you.

Once again, the authors would like to thank the reviewers for their comments. We hope that all queries have been addressed.