Author's response to reviews

Title: A Standard Procedure for Creating a Frailty Index in a Cohort Study

Authors:

Samuel D Searle (ssearle@dal.ca)
Arnold Mitnitski (Arnold.Mitnitski@dal.ca)
Evelyne A Gahbauer (evelyne.gahbauer@yale.edu)
Thomas M Gill (thomas.gill@yale.edu)
Kenneth Rockwood (Kenneth.Rockwood@dal.ca)

Version: 2 Date: 4 July 2008

Author's response to reviews:

Dear Editor:

Thank you for the opportunity to revise and resubmit the paper. As you will see, we have addressed each of the points raised by the reviewers. We viewed Reviewer 3’s Major Essential contention - that counting all relevant deficits was circular - as an opportunity chiefly to clarify the rationale for the Frailty Index. Otherwise, we have made all changes asked of us.

My colleagues and I are grateful for the careful consideration of our work and look forward to hearing from you about the disposition of the manuscript.

Yours sincerely,

Kenneth Rockwood

Reviewer #1

Overview

Reviewer: “… This is a very valuable contribution and publishing this paper in an open access journal would provide a possibility to many researches to have access to this knowledge.”

Reply: We thank the Reviewer for this comment.

Minor Discretionary Comments on the paper.

Reviewer “The authors understand that there is a problem with definition of cut-points for continuous variables. The authors suggested an elegant method for selecting these cut-points by conditioning continuous indices on an interim frailty index. Although such a method looks informative, the authors may wish also to briefly mention other possible approaches as long as they wish to tabulate the process of construction of the frailty index. …”
Reply: We thank the Reviewer for this helpful comment, which exactly fits the spirit of our inquiry, being the elaboration of a method that others could use.

Following the proposal of Streiner & Norman (now reference, 21; was 25 in the first submission) we have broadly categorized as those based on characteristics of the distribution and those based on judgment. (Methods). We have also addressed this point in the Discussion.

Reviewer: It would be helpful the authors define mortality more precisely. Currently, it is unclear how (for each of the two waves of the survey or not) and why (limited by the follow up?) the 5-year mortality was assessed?

Reply: Thank you. We recognize that 5 year mortality assessment was arbitrary and have changed the methods and results to reflect the 9 year and 7.5 year follow up which is the complete follow up data available at this time. We have now corrected:

“Survival analyses were done using bi-variate and multivariate Cox Regression analysis with the frailty index as the independent variable and age and gender as covariates on each of the two survey waves. The survival calculations were based on the available nine year mortality data from the baseline survey. The 18 month follow up survival calculations were based from the available seven and a half year (from 18 month interview) mortality data.”

We have also addressed this in Table 3, and the Results.

Reviewer: I would suggest clarifying a bit the paragraph starting with “In investigating the upper limits to the frailty index…”. Presumably the authors mean the case when they have considered the 99% sample.

Reply: Thank you. Your interpretation is correct; we have added that in the Results.

Reviewer: When the authors discuss predictive validity of the frailty index in Discussion, I do not understand why they did not refer the studies comparing predictive power of the CHS-based definition of frailty and the frailty-index-based definition performed by their group and by others (Kulminski et al., 2008).


Reply: Thank you for drawing our attention to this excellent paper, which only appeared in print after we had submitted our paper. We have added the reference.

Reviewer #2
Minor Essential Revisions

Methods

Reviewer: Could the authors explain why 0.2 as the cut-point on the interim frailty index?

What happens to the performance if you choose 0.3 or 0.1?

Reply: As discussed immediately above, the issue of cut-points is important, and the selection of any one given point is arbitrary. The 0.2 criterion conforms both to distributional considerations (e.g. at the mid-point of the FI values for people characterized as “pre-frail” by the Cardiovascular Health Study phenotypic definition of frailty, and judgment (is associated with a clinically important increase in the hazard rate (i.e. >2).

Methods:

“In addition, setting the value at 0.3 seems unreasonably high, as this is consistently well into the range of frailty, however, defined (including by an increased hazard) so would be insensitive. Greater sensitivity is obtained at a cut-point of 0.1, but with less specificity.”

Reviewer:

Results

Please list the reason for each of the excluded variables. A few seemed counter-intuitive.

Reply: Thank you for this point. We have added to the Results, as follows:

“Variables were eliminated because they did not meet at least one of the five criteria (unrelated to age and adverse outcome, saturated, or there was already ample representation of the system) or because we had identified 40 variables with which to populate the index. Some potential variables excluded were: Distance walked (up to 20ft.) (saturated), admitted to hospital in the past year (non-age associated), use of a walking device (sufficient variables (i.e. n=40) were already included), walking a quarter mile (already accounted for by two variables), measured blood pressure (sitting and standing) (non-age associated), fractures (non-age associated), Parkinson’s Disease (low prevalence), amputation (non-age associated), liver disease (not present in both surveys), taking medication (controversial in relation to adverse health outcome), light/medium/heavy sports (Unreliable prevalence and age association), measured vision (saturation), and various tests of physical performance (already accounted for in other variables and 40 variables already populating the index), such as finger tap and turning in a complete circle.”

Note: several variables could have been included, but that we had reached our target of 40. There is no scientific reason not to include more – we have constructed an FI of 70 items. On the other hand, a recurring criticism of the work
has been that it seems too complicated to calculate an FI if too many variables are used. Here, as in a few earlier studies, we have selected variables at random (boot-strapping) from a list of eligible variables to make up the FI and have shown that the results are insensitive to the precise composition of the index. We now have spelled this out in the Discussion.

Reviewer: Table 2 has a typo in the heading and RR should be replaced with HR
Discretionary Revisions
Reply: Done. Thanks.
Discretionary Revisions:
Reviewer: The effect of sex on mortality is accentuated by adjustment for frailty and seems less apparent on follow-up. Do the authors have a suggestion why this is so?
Reply:
This is an interesting point. We believe it reflects the fact that while men accumulate fewer deficits than do women, any given level of deficit accumulation is more lethal for them. We have added the following to our discussion.

“One notable result from the Cox analysis is that including the frailty index increased the impact of being male on mortality. This likely reflects the observation from earlier studies that while men accumulate fewer deficits than do women, any given level of deficit accumulation is more lethal for them and at any given age, females seem to be more frail than males [6,11].”

Reviewer #3
Major Compulsory Revisions:
Reviewer: While cross-validation of FI across multiple populations is important, the arguments presented in this paper could be made stronger if the following issues can be appropriately addressed. First, if one of the inclusion criteria for a FI is that “the variables must be deficits associated with health status,” assessing its predictive validity via its association with mortality seems a bit circular in the sense that one’s health status is expected to be associated with mortality risk. Also it would be helpful to clarify the definition of “health status.”
Reply:
Thank you for allowing us to clarify some essential points. As the Reviewer intimates, there are many ways to conceive of health, but for the present purpose of understanding variable risk in health outcomes, anything that gives rise to an adverse health outcome could be seen as threatening health. (This would imply that perfect health represents a state of no risk of adverse outcomes, which seems a reasonable way to conceive of such a theoretical state.) We are not clear that what we are doing is circular; a better way to state what we are up to is
that we are quantifying the risk of adverse health outcomes. (In other words, to quantify the frailty state – recall that everyone agrees that frailty is a risk state).

We have now amended the Discussion to read:

“The relationship between the frailty index and mortality is of interest on several grounds, but here is presented chiefly because it represents a relevant and non-arbitrary test of predictive validity. This is important because predictive validity is one of two types of so-called criterion validation, the other being validation against a so-called “gold standard” [21]. As there is no gold standard for frailty assessment, predictive validation is an important method of validating any approach to frailty operationalization. Note that our intent in checking the ability of the frailty index to predict mortality is validation of the index, rather than developing a mortality prediction index that included frailty. If the frailty index were meant to be a mortality prediction instrument, there might be a rationale for weighting several items, particularly age. One notable result from the Cox analysis is that including the frailty index increased the impact of being male on mortality. This likely reflects the observation from earlier studies that while men accumulate fewer deficits than do women, any given level of deficit accumulation is more lethal for them and at any given age, females seem to be more frail than males [6,11].

In consequence, the reason to evaluate deficits is that each has a small effect, but considering them in totality allows the impact of small effects to accumulate. There is nothing circular about this; on the contrary, it is a more exact estimation of risk compared with evaluating just a few variables. Note too that including all 40 in one variable offers the added advantages of dimensionality reduction, and of studying the behaviour of the frailty index itself (e.g. studying its slope in relation to age; the change of slope in relation to age with changes in deficit accumulation, the presence of a limit).

Reviewer: Second, the methodology behind the proposed re-sampling procedure for studying reproducibility of the FI needs to be justified. For example, the choice of 80% for variable selection appears to be arbitrary as the reproducibility is almost certainly a function of the sampling proportion.

Reply: The bootstrapping was used to evaluate a confidence interval, and speaks to both statistical and clinical considerations: i.e., the results are not dependent on just a few variables. There are no published guidelines for boot-strapping by variables to calculate a confidence interval. We have pioneered the use of this bootstrapping by variables approach, as it reflects our contention – which has been amply verified in several studies – that the properties of the Frailty Index are indifferent to its composition.

We used 80% of the variables as 80% corresponds to 32 variables. In principle we can use fewer, say 75% which corresponds to 30 variables, the lowest number of variables we would suggest should be included in a frailty index (see
Methods). Using 75% we obtained 95% confidence intervals for the slope of 0.012-0.028 (baseline) and 0.019-0.033 (18 month follow up).

Reviewer: Finally, the purpose of using the 18-month assessment was not clearly stated.

Reply: It was the planned follow-up of the cohort (as presented in the methods), which aimed to study precipitating events for relevant outcomes.

Reviewer: It is also surprising that those remained in the study at the 18-month visit were frailer because I would expect that those who were less healthy were more likely to drop out. More importantly, what to make of the difference seen in the rates of deficit accumulation between the baseline and follow-up? Was it simply due to sampling error as suggested by the overlapping confidence intervals or meaningful between-person heterogeneity? If it’s the latter, what would be the clinical implications?

Reply: Recall that the 95% confidence intervals for the slopes of the lines that related deficit accumulation are 0.14-0.26 (mean 0.2) and 0.020-0.032 (mean 0.26) indicating that, on average, people accumulated more deficits after 18 months, as the Reviewer points out. The Reviewer expresses surprise that the effect of surviving to 18 months did not mean that the cohort became healthier on average.

There are several points to sort through here. The essential bottom line, however, is that even differential mortality (with survival of the fittest) is not enough to overcome the effect of ageing: in general, the older a person is, the more likely they are both to die, and to have something wrong with them. (Clearly, they are not unrelated.) By quantifying frailty in relation to what things people have wrong with them, the effect of ageing is quantified.

Now, as it turns out, a higher slope of the line that relates age to the number of things people have wrong with them does not necessarily mean that they are frailer; a group with a steeper slope to the age-deficit line might be less frail than a group with a less steep slope. What counts is the absolute level of frailty. (Going back to the line relating frailty/deficit accumulation to age, it is clear that it is the intercept of that line which is important in determining how frail a person is, not the slope of the line.) According to the Reliability Theory of Ageing, loss of the slope of the line that relates hazard to age reflects loss of redundancy of a system (Gavrilov LA, Gavrilova NS. Reliability theory of aging and longevity. In Masoro EJ, Austad SN (eds) Handbook of the Biology of Aging, 6th edition, New York: Academic Press, Ch. 1., esp. Figure 1.8). Briefly, the reliability theory of ageing appears to provide a considerable and considerably useful apparatus for further work on the properties of the Frailty Index, but is outside the scope of the current inquiry.

Minor Essential Revisions:

Reviewer: Results section, 1st paragraph, last line: “Each of these variables were
eliminated because they … failed to cover more than one system…” First, change “were” to “was.” Second, the 4th criterion says that the selected deficits “that make up a frailty index must cover a range of systems.” It does not require that each variable has to cover more than one system.

Reply: Thank you. We have corrected and clarified this issue in the first paragraph of the results section.

Reviewer: Results section, 2nd paragraph, last line: change “cut-points offs” to “cut-offs” Table 3 title: change “RR” to “HR”

Reply: Each done. Thank you.