Author’s response to reviews

Title: Reliability, validity and administrative burden of the community reintegration of injured service members computer adaptive test (CRIS-CAT)

Authors:

Linda Resnik (Linda_Resnik@Brown.edu)
Matthew Borgia (Matthew.Borgia@va.gov)
Pengsheng Ni (psni@bu.edu)
Paul Pirraglia (Paul.Pirralgia@va.gov)
Alan Jette (Ajette@bu.edu)

Version: 5 Date: 3 May 2012

Author’s response to reviews: see over
Dear Editor,

Please find a point by point response to Reviewer 2’s comments on the first revision of this manuscript. We have followed your instructions to focus on the statistical issues raised by the reviewer. We believe that the manuscript has been substantially improved as a result. Thank-you for your consideration of this revision.

**Major Compulsory Revisions**

*The title must be changed. Many clinicians use the term 'Psychometric' incorrectly. Psychometrics is about measurement (see http://en.wikipedia.org/wiki/Psychometrics), and you are obviously interested in prediction (outcomes, change, etc.). Please change 'psychometric evaluation' into something like 'external validation' or 'predictive utility'. Maybe look at McDonald’s 1999 book (Test Theory) for a good term.*

The title has been revised. The new title is, “Reliability, validity and administrative burden of the community reintegration of injured service members computer adaptive test (CRIS-CAT)”

*The article does not read easily. Please provide more structure to it. Explain why you do things; explain the results to the audience instead of solely referring to the table. In your discussion, please discuss the outcomes in the light of your research goals. What can we learn from your study? Try to arrange the sections of your discussion according to content.*

Per the Editor’s advice, in this revision we have focused on responding to the statistical concerns rather than reorganizing and revising the text. We do believe that the results are summarized in the text and are reluctant to repeat information from the table within the text. However, we have augmented the Discussion section considerably and have rearranged several paragraph to improve the organization by content area. We’ve used the following sub-headings to make the organization of the Discussion section clearer: Reliability and Respondent Burden, Concurrent and Known Groups Validity, Predictive Validity, Limitations

*The authors did not deal with the missing values point I made. They are 'unaware of methods for dealing with missing values in small longitudinal studies'. Instead they used complete cases. That this is very problematic is acknowledged in the discussion, and very obvious from Table 2. I would suggest that they multiply impute their data, and use relevant covariates (such as those in the table) in the imputation model to deal with this self selection. Under MAR the assumptions concerning the missingness are less strict then under MCAR (which is done here).*

The Reviewer’s comments speak to the analysis of predictive validity of the CRIS-CAT scales in the longitudinal cohort study. In response to these comments we explored methods to impute the outcomes of SF 12 PCS and MCS scores for the 73 Veterans whose PCS and MCS data was missing at Visit 2 (those lost to follow-up). We then conducted a sensitivity analysis to examine the impact of including cases with imputed missing data by modeling PCS and separately MCS scores using all the covariates in Table 2 as well as the Visit 1 CRIS CAT (extent, perceived, satisfaction), PCS and MCS...
scores. We used STATA 11.1’s mi commands to multiply impute 10 values for each of the SF-12 scores, based on regressions of the covariates and examined the combined results with the mi estimate command.(http://www.stata.com/stata11/mi.html)

These analyses showed that the statistically significant results of the linear regressions predicting PCS were similar for models with and without the imputed outcomes data. The statistically significant results of the linear regressions predicting MCS were similar for the CRIS-CAT Extent and Perceived Scale. However, the relationship between the CRIS-CAT satisfaction score and the MCS was no longer statistically significant (P=0.077) in the model with the imputed MCS scores.

Our examination of predictive validity included outcomes other than the MCS and PCS, and it is unclear to us whether or not the Reviewer is suggesting that we impute these other outcomes as well: self-reported change in marital status, employment status, and housing stability. We, like many other researchers, are uncomfortable imputing primary dependent variables, particularly categorical variables where we have little variance. We also believe that it would be confusing to the reader if we were to impute some of our outcomes (for example the MCS and PCS), but not others.

Given the results of our exploration of imputed MCS and PCS scores and our concerns about imputing multiple categorical dependent variables, we respectfully disagree with the Reviewer’s suggestion to impute the missing data in our primary analyses. As the Reviewer points out, we already recognize that loss to follow-up limits the generalizability of our results to the entire Veteran population. However, we do not believe that this diminishes from an overall finding of predictive validity for our sample. We have augmented this section of the discussion even more, indicating that we did conduct a sensitivity analysis by imputing missing MCS and PCS data and that our results are largely the same, but not identical. However, further studies are needed to confirm our findings related to change in marital status, employment status, and housing stability.

*I don't think the authors dealt sufficiently with my points concerning statistical concepts. The equations are a mixture of symbols and text, which makes it very messy (are they at all necessary to be presented?). Please provide a better and more extensive introduction of all statistical and psychometric terms. Under what model are things obtained? We edited the equations by separating the text from the formula and explaining the calculation of the standard error. Furthermore, we have defined the term ICC. The text already describes MDC. We would be pleased to describe other terms, if the Editor so advises, but we question the need to describe all common statistical and psychometric terminology: i.e. Pearson product correlations, concurrent validity etc. for this audience.

Standard errors of what parameter?! We assume that this question refers to the text on p. 14. We have clarified that this refers to the standard error of the person score estimators.
*Likewise, please say something about latent variables earlier on in the paper. The first reference to them is on page 14, where it just suddenly appears. Thank-you for this suggestion. We have made it clearer that community reintegration is a latent trait by editing p. 6 to read, “The Community Re-integration of Service Members (CRIS), a measure of community reintegration developed to address this gap, measures the latent trait of community reintegration through the assessment of participation in life roles as defined by the International Classification of Functioning, Disability and Health (ICF).”

*The end of the introduction does not all justify this study! The main point should be that now that you have a psychometrically sound CRIS-CAT (which has a small respondent burden), is it useful in practice (or as useful as the full CRIS)? To explore this, you study if you can predict all kinds of (external) measures which are relevant for the clinical field (change, outcomes, concurrent measures etc.). If you can, the CRIS-CAT can be used in the field to select subjects who are at a high risk of showing or developing pathology! This should be your main selling point. Please extend your intro.

We’ve added to the introduction by explaining that, “This type of information would be useful for monitoring of Veteran functioning and for targeting treatment to Veterans at risk for adverse outcomes.”

*The structure of the paper is very awkward (methods per sub-study, results per sub-study). Better describe all three sub-studies separately (Study 1: methods, results, small conclusion; Study 2, methods, etc.)

We reorganized the first revision of the manuscript so that the methods and results sections of each of the sub-studies were described separately and added a new Figure that shows the relationship of the 3 sub-studies. The other two Reviewers were satisfied with this reorganization and we believe that it works well. We are maintaining the structure in this revision, guided by the Editor’s advice to focus this response on the statistical concerns.

*I truly don’t understand what you did in the ‘likelihood of change in marital status analysis’! As I understand it, having two times three values, would give nine possible combinations. Why would this give three levels of change? How would a multinomial regression model incorporate this change?

*Your prediction of change in marital status cannot be assessed if 121 of 135 do not change. Your variance is almost zero; therefore a trivial outcome!

P. 12 explains that “Change in Marital Status at Visit 2 was categorized into three groups: newly married, no longer married and unchanged. Newly married status was defined as change from any non-married group at baseline to married at 1 year follow-up. The no longer married status was defined as change from married at baseline to any non-married group at 1 year follow-up. We classified any subjects who remained married or remained in the non-married groups as having an unchanged marital status.”

We are aware that we lack variance in this outcome and therefore we don’t show the results. We still believe it is important to report on this analysis, but acknowledge that it
is limited. We added to the discussion section by stating, “In the current study, we did not find a relationship between CRIS-CAT scores and 1-year change in marital status, housing, or employment, as we had hypothesized. However, these findings might be attributable to the very small number of Veterans who had a change in marital status, the relatively short time frame of our follow-up and small numbers of veterans who had changes in these domains during the follow-up period.”

Minor Essential Revisions
*There are still many typos (to assess to assess?), and inconsistencies [(within brackets Perceived mn 56.6 ±11.5) and outside brackets Satisfaction (mn 54.9 ±11.0), what is mn?].
We have proofread the paper and corrected issues with copy-editing and bracketing.

*If all p-values in a table are equal, say so in a note an delete column!
Thank-you for this suggestion, we edited Table 3 accordingly.

*What is ability (p.6) in the context of community registration? It is based on self-report, and therefore incomparable with proficiency testing interpretations!
We removed the term “ability” from this sentence which now reads, ” Computer Adaptive Test (CAT) methodologies use a computer interface to select and administer items tailored to the unique level of the respondent.”

*Page 7, ‘reliability equal to 0.9. IRT has no single reliability outcomes. It has conditional reliability (SE of latent trait, and related Item Info) which is sometimes transformed to a more intuitive 0-1 scale (marginal reliability?). Please mention this!
We agree with the Reviewer that IRT has no single reliability estimate. In this manuscript we report on the conditional reliabilities derived from the SE of each latent trait. The 0.9 single reliability score refers to reliability when using our stopping rules. We have edited the text to read, “The CAT stopping rules were set as a minimum of 10 items and a maximum of 20 items and a SE of the latent trait set to 0.32. Reliability of the CAT under these conditions was equal to 0.”

*Page 7, How can simulations tell something unidimensionality, and model fit? An IRT model is usually fit before CAT construction! Please change in a proper fashion!
The Reviewer is correct. Examination of unidimensionality was done through IRT analyses which proceeded CAT construction. We added a statement to that effect prior to the sentence that discussed the results of data simulations on p. 7, and edited the statement in question.

*Page 8, please change term 'psychometric characteristics' into something appropriate (see above).
The term psychometric was replaced.

*Page 9, acronyms of questionnaires, please give full name at first reference for
all scales.
We have spelled out the full names for the Craig Handicapped Assessment Reporting Technique (CHART), and the Quality of Life (QOL) measure. To the best of our knowledge the SF-36 V does not have another title.

*Page 14, true variance should be true score variance; what is THE latent variable. The latent trait under what IRT model (the one you used, so GRM of PCM or other?).
Thank-you for this comment. Our assumptions referred to the latent trait under the partial credit model. We edited the text accordingly.

*Same page, person standard error. I guess this should be SE of the persons estimated latent trait? As above, reliability as you use it is something that is transformed from the real thing (SE?) to provide non-technical users with.
The SE is of the person score estimators.

Explain this!
*For all kinds of technical measures you refer to papers of applied researchers using them (e.g., references 25, 26, 21, 27, 28). Instead, please refer to the original methodological/statistical sources/references where these measures were developed. Because it is uncertain if statistics were assessed properly in applied papers (and we therefore do not know if they were used properly), it is a good habit to refer to the original sources.
We added additional original references to technical methods and omitted references 25, 26, 27 and 28. We are retaining several references. Reference 21 refers to the CHART measure and was written by the developer. Reference 26 cites previous research that supports our hypotheses.

*Page 19 and Table 6. What do the MDC values tell us? Please provide explanation.
The discussion of MDC values was moved from p. 16 to the discussion section, pgs 19-20.

*Please, for each table, state for which study (1, 2 or 3) it is relevant.
Thank-you for this suggestion. We amended the titles for several tables where this information was missing. We prefer to use the names “Field Study”, “Administration Study” and “Longitudinal Cohort” rather than numbers 1,2 or 3 to be consistent with terminology used Figure 1.

Table 1. Descriptive characteristics of subjects in the field study by group

Table 2: Characteristics of participants in the longitudinal cohort study and Administration study

Table 3: Concurrent and Discriminant Validity of CRIS-CAT scales: Pearson Product Correlations, data from the field study
Table 4. Key Outcomes at Visit 2: Longitudinal cohort study

Table 5 Linear Regression Predicting SF-12 scores at Visit 2: Longitudinal cohort study

Table 6: Administration Study: Summary of raw scores, number of items used, ICCs and MDC value