Reviewer's report

Title: A retrospective cohort study of U.S. service members returning from Afghanistan and Iraq: Is physical health worsening over time?

Version: 4 Date: 6 November 2012

Reviewer: Andrea Finlay

Reviewer's report:

I appreciate the author’s attention to both reviewers’ comments. Notable improvements have been made to the manuscript including reanalysis with 40 sets of imputed data, inclusion of gender as a covariate and further details on the samples. However, I continue to have extensive methodological concerns which I will outline below.

Major Compulsory Revisions

1. A longitudinal community sample is included to enhance the generalizability of a cross-sectional clinic sample. However, there are notable differences in the samples (e.g., age, military branch, PCS score and subscale scores) that indicate that these samples would be better used in separate studies.

2. On a related note, the author did not find a significant difference between immediate post-deployment PCS and one-year post-deployment PCS suggesting that the results from the clinical sample are not generalizable to the community sample.

3. Causal language is still present throughout the manuscript. For example, “OEF/OIF veterans evaluated at our clinic deteriorated even after adjusting for PTSD”, “PCS scores significantly decreased with increasing time post-deployment,” and “veterans in our clinical sample report substantially poorer physical health over time.” This point was mentioned by both reviewers and was not adequately addressed by the authors.

4. The analyses conducted are still unclear. Why were ANCOVA analyses used instead of OLS with time from deployment as a continuous variable? A veteran who was deployed 365 days ago would be in a different group than a veteran deployed 366 days ago. I do not find the argument outlined in the letter convincing and would appreciate some theoretical or empirical reasons why veterans were categorized by year. Furthermore, previous work has indicated that dichotomizing variables when there are not natural groups present (e.g., gender, treatment/control) introduces bias and the same issues are present in this study when dividing veterans into four groups (e.g., MacCallum, Zhang, Preacher, & Rucker, 2002; Royston, Altman, & Sauerbrei, 2006).

5. At what time point was the PCL available for the community sample? Why wasn’t it used as a control at one time point? Was PCL used in the MI model as an auxiliary variable as PTSD has links with physical health?
6. In the Statistical Analysis section I expect it should clearly state from start to finish which analyses were conducted. The clinical sample should indicate that MI was not conducted because missing data was less than 5%. Then it should state which tests were conducted to compare groups. Why was a chi-squared test conducted to test differences in age when a t-test would be more appropriate for continuous variables?

7. Why was a combination of PTSD, age, and gender used in the ANCOVA analyses? No explanation for including a three-way interaction was in the Introduction. Has previous research found associations between these variables? Also, why was a three-way interaction conducted and not a two-way interaction of PTSD and gender?

8. I still cannot determine from the Statistical Analysis section or Table 2 what tests were conducted. Was year the variable used for the linear contrasts? Was each covariate tested in a separate model along with each physical health variable? In other words, what does the F-test value for covariate PTSD and variable PCS represent? Was age, gender, etc. not included in this test? Or if they were why are there separate F-test values for each covariate?

9. On a related note, Table 2 is difficult to read and does not conform to standards for ANCOVA tables. Furthermore, it would be helpful to have the beta coefficients in the table and statistical significance indicated with asterisks as is standard in scientific literature. A table reporting all of the results would be helpful to understand what tests were conducted and what the results indicate.

10. Was a Bonferroni correction used in both the cross-sectional samples and longitudinal samples?

11. The Results section indicates that (paired?) sample t-tests were used to examine changes in PCS over time in the community sample. Why are the degrees of freedom different when comparing pre-deployment with immediate post-deployment and pre-deployment with one-year post-deployment? The degrees of freedom were again different when comparing immediate post-deployment with one-year post deployment. The degrees of freedom should be the same since MI was used.

12. The authors state in the Physical functioning in OEF/OIF veterans section that the different between immediate deployment and one-year post deployment approaches significance, but the t-value and p-value indicate that it was significant. Please clarify which is correct.

13. In the Discussion section it would be clearer to state what the normal decline in health during a 2-year period would be for people in their early 30s. The added sentence is more confusing.

14. I cannot adequately assess Discussion/Conclusions section until the analyses are fully described and any methodological issues are resolved. Generally, the Discussion section will need to be revised once the above issues are addressed.

15. Table 1 appears unnecessarily wordy in places. For example, why not list the military branches in the left column and only list the percentages in the remaining
columns?

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.