Reviewer’s report

Title: Monitoring the health of gender minority populations: Cognitive testing of natal sex and gender identity survey items in a U.S. national cohort of young adults

Version: 2
Date: 7 August 2014

Reviewer: Greta Bauer

Reviewer’s report:

This manuscript, entitled “Monitoring the health of gender minority populations: Cognitive testing of natal sex and gender identity survey items in a U.S. national cohort of young adults”, presents a well-written report of an evaluation of both construct validity and cognitive testing for a two-part measure of gender that includes natal biological sex and current gender identity.

As we move toward incorporation of explicit inclusion of trans participants in health research and other population surveys internationally, the work presented here is critical to the design and validity of much future research. This is an innovative use of an existing longitudinal data set with measures at three time-points. Its major limitation is that the sample – those who are young adults, the children of female nurses (primarily white), and who complied with years of follow-up research – are those probably least likely to have problems understanding these questions, or to be offended by them. That this study shows good performance of these measures is a start, and requires that additional research then expand to a broader population, which the authors appropriately note.

This paper represents a strong contribution to the public health literature, with appropriate design and analysis, and clear organization. Nevertheless, I have suggestions for improvement, which I have broken into categories below as requested.

Major Compulsory Revisions

1. While limitations of the study population are discussed on page 16, this paper needs to address the additional limitation that the sample represents only the subgroup who was retained during years of follow-up (1996 to 2010). Only 7,831 of 16,882 original participants (46%) are included in this analysis. Can information be included regarding how these participants differ from the GUTS1 participants who were lost to follow-up or did not complete the relevant measures and were thus not included in this analysis? Without this information, it’s difficult to know to what extent these results apply to the group of young adult children of female nurses, or whether factors that affect retention may also create an even more narrowly circumscribed and potentially non-representative group.

Minor Essential Revisions
2. The second part of the title does not accurately reflect the paper’s focus on construct validity as well as cognitive testing. It could be reworded to begin with “Construct validity and cognitive testing of…” or alternately just “Validity of…”.

3. p. 6, line 20: Please specify the type of correlation statistic used.

4. p. 6, line 22 and p. 7, line 14: Please specify the statistic used for #. I would assume Cronbach’s, but it’s not clear.

5. Explain the decision to categorize variables (e.g. dichotomizing gender nonconformity scores). Why throw out potentially meaningful data (with likely reduction in statistical power) by reducing to categories?

6. Of the 26 participants who are coded as gender minority, the majority (14) did not identify as male, female or transgender. This jumps out, and needs more discussion. While we cannot know how these participants would have responded given only the options of male, female or transgender, it suggests that it may be quite important to include the other option. It may be most important among youth, given both trajectories of identity formation and potential cohort effects regarding rejection of identities and acceptance of broader gender diversity, and it would be interesting to see if this would look quite different among older persons. Given that this group was represented by only 2 of the 9 participants in the cognitive testing, it also remains unclear whether or not this group may contain some who would not in fact be appropriately categorized as gender minority, though the responses of those 2 are reassuring.

Discretionary Revisions

7. Given that there is so little research on measurement of gender in health research, I do have some concerns that readers may extrapolate that this method of measurement is the best for all studies. While I agree it’s probably the best for large multi-use population data sets, individual research projects may require different measures to meet their aims beyond just identifying trans participants. This item includes on domain of biological sex (that assigned at birth, usually based on external genitalia) and one domain of social gender (identity). For some health studies, lived gender will be also be relevant, and for others, different biological domains will be relevant (e.g. hormonal milieu or current anatomy). It would be good to make this point for those designing their own studies, so they are able to use their data and interpret their results. While I dislike self-referencing, I believe I’m still the only one to have written about this with regard to trans items in surveys (Chapter 8 in this book: http://www.cihr-irsc.gc.ca/e/documents/What_a_Difference_Sex_and_Gender_Make-en.pdf). This is really just an extension to methodology of work on domains of sex/gender by multiple people (e.g. Anne Fausto-Sterling, Milton Diamond).

8. It may also be worth commenting on the frequency of being transgender within the sub-sample of GUTS1. There were 26 participants out of 7,831, which gives a prevalence of 0.3%. This is similar to the estimates the authors and others have produced for adults in two U.S. states, which is interesting considering the
young age of this group.

9. On page 11, in the paragraph on response options, it may be worth noting that while the participant thought male and female options alone were relevant because they reflect the options on birth certificates, that in some jurisdictions there are third options such as “undetermined” that are assigned (e.g. in Ontario, Canada). I am not aware of any good reviews of current birth certificate options across jurisdictions, but researchers may want to ensure their options accurately reflect birth certificate options, which may or may not be from the jurisdiction in which participants currently live.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.