Author’s response to reviews


Authors:

Adam C Carle (adam.carle@unf.edu)

Version: 2 Date: 5 December 2008

Author’s response to reviews: see over
December 5th, 2008

Re: MS: 9643275022221554

Melissa Norton, MD Editor-In-Chief
BMC Public Health

Dear Dr. Norton,

Attached please find a copy of my revised manuscript now entitled “Assessing the Adequacy of Self-Reported Alcohol Abuse Measurement across Time and Ethnicity: Cross-cultural Validity across Hispanics and Caucasians in 1992, Invalidity in 2001-2002.” I am submitting this revised manuscript for you to consider publishing in BMC Public Health. Both reviewers felt the original manuscript addressed important issues. For example, reviewer 1 felt the manuscript fills a “gap [in] our knowledge regarding the concept and measurement of alcohol [abuse],” while Reviewer 2 indicated that the manuscript “raises an important question with respect to the cross-cultural validity of self-reported alcohol abuse symptomatology.” Reviewer 2 also indicated that “the methodology … was appropriate for the intended purpose, adequately described and appropriately implemented.” However, some concerns tempered the reviewers’ enthusiasm. I appreciated these critiques, felt they would strengthen the study, and incorporated all of them into this revision. Let me detail these changes:

**Reviewer 1:**

1. The author referred to Models 1, 2, 3a, 3b and 4 on p.8.. [and] again…on p. 9. … the author referred to these models to illustrate the steps involved in examining measurement invariance the readers may find it confusing especially no actual models are presented.

To address the concern that readers will find the presentation confusing, I utilized the reviewer’s suggestion below. I now present Figures 1-8 (they begin on page 25) to pictorially show what each measurement model tested. I discuss the figures when describing measurement bias analyses generally. For example, on page 7, I include the following:

“Figure 1 presents a visual representation of this measurement model. The solid black circle represents the latent variable, here alcohol abuse. The small circles represent the continuous latent response variates underlying the dichotomous items (represented by the squares). The arrows from the solid black circle to the smaller circles represent the loadings. The arrows
from the small circles to the squares represent the thresholds. Finally, the arrows pointing only to the squares represent the uniquenesses.”

And, on page 8, I now note:

“Visually, for the least restricted model in the 1992 data, this would mean fitting a model like that presented in Figure 1 for each group and allowing the measurement parameters to vary across the groups (excepting those constrained to equality for statistical identification, see below). Thus, in Figure 1, dashed lines represent measurement parameters allowed to vary across groups and the solid lines represent measurement parameters constrained to equality across the groups. As Figure 2 shows, the model constraining the loadings across the groups has solid black lines from the latent variable (solid circle) to the latent response variates, indicating that the loadings have been constrained to equality. Figures 3 and 4 continue the visual representation for the 1992 data. Figures 5 through 8 illustrate the models for the 2001-2002 data."

I also refer to the appropriate figure in the text to provide the reader with a visual image of what the model tested in the results section (cf. p. 9).

Finally, to further reduce confusion, I use a new numbering sequence. I now continue the model numbering sequence when I examine the 2001-2002 models (e.g., I start with Model 5) rather than repeating the numbering (e.g., starting over with Model 1). For example, in the original manuscript, I labeled the 1992 baseline model as Model 1. I also labeled the 2001-2002 baseline model as Model 1. I now label the 2001-2002 baseline model as Model 5 to clarify that it differs from the 1992 model. This occurs on page 9.

2. The author concluded that in 2002-2002 there existed measurement bias which underestimated the true level of alcohol abuse among Hispanics (based on the impactful measurement bias for seven of ten items). Is the author able to provide estimates of the true level of alcohol abuse among Hispanics, as well as the true difference when compared to the non-Hispanic Caucasians based on the 10-item-set being used to assess the level of alcohol abuse in the United States?

The reviewer notes a valid concern and Reviewer 2 raised a similar issue. The NLAES and NESARC rely on self-report. Thus, the differences observed in this study may reflect differences in self-report rather than differences based on diagnoses or some other external criterion. Measurement bias work typically selects a reference group (non-Hispanic Caucasians here) and interprets the findings relative to that group. Given that the alcohol abuse concept was developed in among the majority population, I chose non-Hispanic Caucasians as the reference group. However, lacking a gold standard external criterion, the reviewer correctly raises the concern that these findings address self-reported levels of alcohol abuse. Likewise, without lack of an external gold standard criterion, the findings may reflect over-reporting for non-Hispanic Caucasians (as Reviewer 2 also noted). I address these issues in several ways in the revision.
First, I changed the title. The title now includes “Self-Reported” and reads “Assessing the Adequacy of Self-Reported Alcohol Abuse Measurement across Time and Ethnicity” rather than “Assessing the Adequacy of Alcohol Abuse Measurement across Time and Ethnicity.”

Second, throughout the manuscript, I have added “self-reported” to clarify that self-reports may not match actual statuses. For example, the third to last sentence of the abstract now reads “Bias underestimated Hispanics’ self-reported alcohol abuse levels” rather than “Bias underestimated Hispanics’ true alcohol abuse levels.” Similarly, page 4 now reads “true self-reported” rather than just “true.” I did this throughout the manuscript.

Additionally, throughout the revision I offer a more qualified interpretation and emphasize that the differences reflect bias relative to non-Hispanic Caucasians as the reference group. For example, in the last sentence of the results section in the abstract and in the third to last sentence of the abstract, I added “relative to Caucasians.” Similarly, on page 4, I changed “among Hispanics” to “across non-Hispanic Caucasians and Hispanics.” I also made a parallel change on page 12, changing “… how valid are .. estimates … among Hispanics” to “…how valid are … estimates … across Hispanics and non-Hispanic Caucasians?” I also modified the concluding question to reflect this qualification. The question now reads “Summarily, how well does the field currently measure and estimate alcohol abuse across non-Hispanic Caucasians and Hispanics?” rather than “among Hispanics.” I made related changes on pages 11, 12, 13, 14, and 15.

Finally, I specifically added the reviewers’ (both Reviewer 1 and 2’s) concern as limits in the discussion section. Thus, on page 14, I now include the following:

“Finally, these data represent self-reports and may not reflect actual experiences. Without an external gold standard criterion, it remains unclear the extent to which self-reports differ from actual experiences. Additionally, this leaves open the possibility that these questions provide more accurate measurement for Hispanics and poorer measurement for non-Hispanic Caucasians. In other words, without a gold standard, it is possible estimates over-report non-Hispanic Caucasians’ alcohol abuse levels rather than under-reporting Hispanics’ alcohol abuse. However, given the development of alcohol abuse among the majority non-Hispanic Caucasian population, it seemed reasonable to use non-Hispanic Caucasians as the reference group, as much of the measurement research does.”

I added language on page 14 to explicitly address that future research should address these issues:

“Finally, future research should collect additional data and use an external criterion and examine the extent to which self-reports correspond to the external criterion across non-Hispanic Caucasians and Hispanics. This would clarify whether these findings reflect under-reporting for Hispanics or under-reporting for non-Hispanic Caucasians.”

3. This reviewer highly recommends the use of diagrams to illustrate and clarify the actual modeling procedures to facilitate understanding.
As described earlier, I now use Figures 1-8 to illustrate the models.

4. The authors need to clarify what RMSEA, CFI, TLI, and NCI stand for.

I expanded these terms on their first use in the manuscript (see page 8) and give the preferred values suggested by Hu and Bentler (1999). This section now reads, “...root mean square error of approximation (RMSEA) values less than 0.05; comparative fit index (CFI), Tucker-Lewis Index (TLI), Gamma Hat values greater than 0.95; and McDonald’s noncentrality index (NCI) values greater than 0.90.”

5. In examining the impact of item-level measurement bias on the overall estimates the sensitivity analyses ... present the various statistics... (3rd and 4th paragraphs, p.10) in a tabular format...

Table 3 (page 24) fully summarizes the statistics included in page 11’s impact section.

6. [Move] the sub-section on analytic strategy (p. 6-8) in Results .. to the end of the Methods section.

I moved this section (now 2.4) to the end of the Methods section. This now begin on page 6.

Reviewer 2:
1. ... It is unclear to me whether these items represent the total of the items in the AUDADIS that are considered to be relevant to this diagnosis, or whether the author had selected these items himself ....[Also, the] dichotomous ... response [nature]... should be made clear in the Methods section where the author describes the measures ...

I now clarify that I used the full set of items in the AUDADIS operationalizing alcohol abuse at each time period. I also explicitly state the dichotomous nature of the items in this section.

Page 6 now reads, “The ... AUDADIS used in the NLAES uses a total of 6 dichotomous items to operationalize alcohol abuse criteria. The AUDADIS provides a fully structured diagnostic interview schedule that includes modules to measure alcohol and drug use ... and family history of alcohol and drug use disorders. .... I used all 6 items operationalizing alcohol abuse criteria.”

And, also on page 6, the manuscript now reads, “The ... AUDADIS-IV used in the NESARC uses a total of 10 dichotomous items to operationalize alcohol abuse criteria....I used all 10 items operationalizing alcohol abuse criteria.”

2. .....Some additional descriptive statistics are needed... either a) means and standard deviations ... or, more appropriately ... b) the proportion of individuals endorsing each abuse symptom. This can be ... incorporated into the existing tables ....
Tables 1 and 2 (pp. 22 and 23) now include the proportion of individuals endorsing each symptom.

3. Although the results of the analyses support the possibility of cultural bias for the 2001-2002 scale, subsequent conclusions should be qualified. In particular, a plausible alternative conclusion could be that the 2001-2002 scale accurately assesses Hispanics and over-reports alcohol-related problems in non-Hispanic Caucasian individuals. Without reference to some external criterion of known validity, there is no way to be certain which is true.

The reviewer makes an important point. Reviewer 1 made a similar point. As noted above, I now offer more qualified conclusions throughout the revision.

First, I changed the title. The title now includes “Self-Reported” and read “Assessing the Adequacy of Self-Reported Alcohol Abuse Measurement across Time and Ethnicity” rather than “Assessing the Adequacy of Alcohol Abuse Measurement across Time and Ethnicity.”

Second, throughout the manuscript, I have added “self-reported” to clarify that self-reports may not match actual statuses. For example, the third to last sentence of the abstract now reads “Bias underestimated Hispanics’ self-reported alcohol abuse levels” rather than “Bias underestimated Hispanics’ true alcohol abuse levels.” Similarly, page 4 now reads “true self-reported” rather than just “true.” I did this throughout the manuscript.

Additionally, throughout the revision I offer a more qualified interpretation and emphasize that the differences reflect bias relative to non-Hispanic Caucasians as the reference group. For example, in the last sentence of the results section in the abstract and in the third to last sentence of the abstract, I added “relative to Caucasians.” Similarly, on page 4 I changed “among Hispanics” to “across non-Hispanic Caucasians and Hispanics.” I also made a parallel change on page 12, changing “…how valid are … estimates … among Hispanics” to “…how valid are … estimates … across Hispanics and non-Hispanic Caucasians?” Other examples occur on pages 11, 12, 13, and 14.

Finally, I specifically added the reviewers’ (both Reviewer 1 and 2’s) concern as limits in the discussion section. Thus, on page 14 I now include the following:

“Finally, these data represent self-reports and may not reflect actual experiences. Without an external gold standard criterion, it remains unclear the extent to which self-reports differ from actual experiences. Additionally, this leaves open the possibility that these questions provide more accurate measurement for Hispanics and poorer measurement for non-Hispanic Caucasians. In other words, without a gold standard, it is possible estimates over-report non-Hispanic Caucasians’ alcohol abuse levels rather than under-reporting Hispanics’ alcohol abuse. However, given the development of alcohol abuse among the majority non-Hispanic Caucasian population, it seemed reasonable to use non-Hispanic Caucasians as the reference group, as much of the measurement research does.”
I added language in the last full paragraph on page 14 to explicitly address that future research should address these issues:

“Finally, future research should collect additional data and use an external criterion and examine the extent to which self-reports correspond to the external criterion across non-Hispanic Caucasians and Hispanics. This would clarify whether these findings reflect under-reporting for Hispanics or under-reporting for non-Hispanic Caucasians.”

4. Furthermore, even if the scale under-reports Hispanics’ true abuse levels, the fact that the individual items themselves were not uniformly reflective of this trend suggests a more qualified interpretation. …

I now offer a more qualified interpretation and note that none of my previously offered explanations may fully account for the findings. I now refer to mechanisms generally rather than “causal” mechanisms, and I discuss bias’ “source” (page 13) rather than its cause. I have included the reviewer’s point that measurement differences may accurately reflect legitimate patterns of abuse unrelated to bias. Additionally, I highlight that these mechanisms likely operate “simultaneously” given bias’ non-uniform presentation. These changes occur on page 13, and read:

“A number of mechanisms may simultaneously result in this bias, particularly given bias’ non-uniform distribution (e.g., some items were more difficult to endorse, others were easier). For example, research notes cultural differences in social desirability and the extent to which Hispanics see psychiatric symptoms as undesirable.[48] Hui and Triandis[83] note that cultural differences in sincerity may influence Hispanic responses. Language skills and socioeconomic variability may also differentially affect responses across these groups.[48] Additionally, the findings may not represent error, per se, but rather accurately reflect fundamental differences in alcohol abuse patterns across non-Hispanic Caucasians and Hispanics. Each of these influences may lead to measurement bias and none likely fully account for the findings. Future research should seek to elucidate what leads to these differences.”

Finally, on page 13, I removed “Cultural differences in language use may result in the bias rather than problems with the behavioral exemplars and criteria themselves. Thus, newly worded items presented identically across cultures might eliminate measurement bias while still capturing the decisive behavioral criteria and eliminate the need for culturally specific criteria or items.” Given the reviewer’s point that differences may reflect real differences in alcohol abuse.

5. … the author should use a more generic term when referring to fit indices in a general sense.

I now follow Hu and Bentler (1999) and use the general term “fit index” when referring to fit indices generally (cf. the last paragraph of the Methods section on page 8 and the first sentence on page 9).
I believe these changes address the reviewers’ concerns. If you or they have any questions about these revisions, my response, or require any further information, please contact me at adam.carle@unf.edu or 904-620-3573. Thank you for your continued consideration.

Sincerely,

[Signature]

Adam C. Carle, PhD