Reviewer’s report

Title: A Structural Equation Model of Perceived and Internalized Stigma, Depression, and Suicidal Status among People Living with HIV/AIDS

Version: 0 Date: 15 Sep 2017

Reviewer: Torsten Neilands

Reviewer’s report:

Review of PUBH-D-17-01468

A Structural Equation Model of Perceived and Internalized Stigma, Depression, and Suicidal Status among People Living with HIV/AIDS

General Comments:

This interesting manuscript investigates a set of important research questions involving the amount of perceived and internalized stigma, depression, and suicidal ideation and attempts among a medium-sized clinic-based sample of urban HIV-infected individuals located in China. The research questions are significant and timely and the descriptive data the authors present are interesting and important. The authors also use structural equation modeling (SEM) to investigate the direct and indirect associations among stigma, depression, and suicide variables. Overall, this methodological approach is consistent with other published works (e.g., Choi et al's "The Effects of Sexism, Psychological Distress, and Difficult Sexual Situations on U.S. Women's Sexual Risk Behaviors", AIDS Education and Prevention, 23(5), 397-411, 2011). Enthusiasm for the current manuscript is reduced somewhat by a number of weaknesses, many of which are minor and listed below. Several more significant issues worth addressing in a more complete and nuanced way include: a need for greater clarity in how the variables are operationalized for the various analyses conducted, the basis for the two-factor solution for the PIS, the decision to exclude cases with incomplete data listwise, a need for a more complete description of the informed consent and interviewing process and what steps were taken to mitigate against socially desirable responses to the survey questions, more details on how many participants refused to participate (refusal rate), the potential biases inherent in a clinic-based sample, and the limitations of cross-sectional mediation analyses.

Also, if the authors' goals are dual-fold to first report on descriptive statistics and then as a second goal to fit their SEM to examine the relationships between stigma, depression, and suicide variables, it could be helpful for the reader for the authors to more explicitly organize the data analysis description in the Methods and the findings presented in the Results under those two general headers. In the current version of the manuscript, the descriptive findings get dominated by the SEM results. This is a shame because the descriptive findings are very important and interesting and the cross-sectional nature of the data limit the conclusions which can be drawn from the SEM, especially with respect to mediation.
Addressing these and the other issues listed below would result in a stronger, clearer publication that would be accessible to a wider audience and therefore be more likely to make a significant public health impact. Finally, I'd recommend the authors seek a careful line editing for grammar and language. I noticed multiple instances of missing articles such as "the" and "a" and occasional misspellings (e.g., "Spearmen correlations" instead of "Spearman correlations"); attention to grammatical issues will also make the paper more accessible and appealing to a wider audience.

Specific Comments

1. Abstract: The SEM global model fit is described as "good", but the abstract Methods don't say what criteria were used to determine satisfactory global model fit and the abstract Results don't report the global model fit, so the reader cannot evaluate from the abstract whether the model fit is satisfactory (minor weakness).

2. Methods, Study Site: I noted that the participants were recruited from both inpatient and outpatient departments. Can the authors please add to the demographics table the proportion who are inpatients vs. outpatients? Can the authors also justify pooling inpatient and outpatient populations together or state as a limitation that the data from these two presumably different groups of patients are pooled and invite further research with larger samples of both groups to enable testing the equality of their factor structures and structural relationships? (minor weakness)

3. Methods, Participants and Sampling: The authors noted that those who were mentally ill or unable to finish the questionnaire were excluded. How many were excluded based on each of these two criteria? Also, by "mentally ill", do the authors mean mental illness of a severity such that the patients would be unable to complete the survey instrument? After all, suicidal ideation and certainly suicide attempts could be indicators of mental illness in general, so it's important to clarify what type of mental illness would exclude someone from this study (minor weakness).

4. Methods, Participants and Sampling: The authors indicate that participants were recruited through direct outreach in the clinic waiting room. Can the authors please add information on how many potential participants refused participation? If that number is not trivially small, there could be bias due to a selection effect. That is, the participants who feel most stigmatized may elect to refuse to participate in the study, which would limit the generalizability of the results. (moderate weakness)

5. Methods, Participants and Sampling: Relatedly, the authors do not describe how the data were collected (mode of survey administration). If interviewer administered, what steps were taken to reduce social desirability bias - respondents may have felt less comfortable admitting to suicidal ideation or suicidal behaviors or depression to another person than via a computerized interview instrument. (moderate weakness).
6. Methods and Sampling: While ethics training for the interviewers is mentioned, informed consent is not discussed (moderate weakness).

7. Methods, Participants and Sampling: The authors note that 39 out of 450 participants had incomplete data (approximately 8.7%), so only the subsample of 411 participants who had complete data were included in analyses. Recommended best practice for addressing missing data is to include cases with incomplete data in analyses to maximize generalizability and statistical power when the amount of missing data exceeds approximately 5% of the sample. I therefore recommend that the authors re-fit their models using a modern missing data handling SEM methodology such as full-information maximum likelihood estimation (FIML) or multiple imputation (MI). A helpful reference for these methods is John Graham's Missing Data Analysis: Making it Work in the Real World in the Annual Review of Psychology, 60, 549-576, 2009. An excellent textbook treatment of Missing Data and modern methods to address it is Craig Enders' 2010 text Applied Missing Data Analysis (moderate weakness). Mplus supports FIML via the MLR estimator, so for their measurement model all they would need to do is include the cases with both complete and incomplete data and re-run their measurement model. For the structural model estimated via WLSM, Mplus supports multiple imputation (MI), so the authors could try re-running their full SEM using the Mplus MI feature; if the results do not differ substantively, they could indicate that they performed MI as a sensitivity analysis and got similar results and so report the current results based on listwise/casewise deletion in the paper - or simply update the paper's tables and results narrative to report the more optimal MI-based results.

8. Methods, Socio-demographic characteristics: Please provide the values for categorical variables and units for continuous variables in parentheses in this section. E.g., instead of "age, gender", please provide, "age (in years), gender (0 = male; 1 = female)." Also, the "and etc." at the end of this subsection's sentence is unclear. I recommend listing each demographic variable explicitly (moderate weakness).

9. Methods, PIS scale: I am confused about the use of exploratory factor analysis (EFA) in this context: if the PIS has been validated already in China, as stated in this section, presumably those validations included a recommended factor structure, so why didn't the authors just use the existing factor structure? On the other hand, if the factor structure of the PIS was not explored as part of the previous validations of the PIS in China, then I would prefer to see much more detail about the factor analysis to strengthen the argument that this scale is indeed a two-factor scale. For instance, what software program and estimation method (keeping in mind that with 4 or fewer categories, a categorical data estimation method is recommended per Rhemtulla et al's 2012 simulation study in Psychological Methods, 17(3), 354-373) were used to extract the two factors? What criteria were used to determine the number of common factors? (moderate to major weakness)

10. Methods, PIS scale: The authors provide no explanation or justification for their decision to trichotomize the continuous PIS scores. Categorization of continuous variables without justification may be suboptimal as described in R.C. MacCallum et al's paper "On the

11. Methods, Depression scale: At the end of this subsection, it is stated that PLWH who have depression scores of less than 16 were considered as having depressive symptoms. That implies that the depression variable may have been dichotomized, yet in the data analysis section depression is described as if it is a continuous scale score. I found this seeming inconsistency to be confusing. The same is true for PIS: as noted in comment #10 above, it is mentioned that the PIS scale was trichotomized, but then its subscales appear to be used as continuous in the SEM. To prevent other readers from having similar confusions, it would be very helpful to add a sentence at the end of each measurement description paragraph that states how each variable will be treated in the data analyses. E.g., For depression, you might add something like, "For purposes of the analyses described below, depressive symptoms was measured by the four constructs (depressed affect, positive affect, somatic and retarded activity, and interpersonal problems) with each treated as a continuous subscale score." By the way, I assume positive affect was reverse-scored since its factor loading shown in Figure 1 is positive and the factor loading for the other three negative mental constructs are also positive? If so, that should be noted in the paragraph where the depression subscales are described. (moderate weakness)

12. Data analysis, item parcels: Item parceling remains a controversial subject in SEM, with some SEM experts recommending it under some circumstances and others counseling against this practice entirely. In particular, Marsh et al (pp. 257-284) and Little et al. (pp. 285-300) in the same 2013 issue of Psychological Methods (v. 18, no. 3) may be helpful references to consult for best practices for item parceling. At the very least, I would urge the authors to provide much more detail on what recommended best practices they followed to construct item parcels, how they constructed their item parcels, and to provide a strong justification for using parcels over traditional independent items CFA. (major weakness). After examining the model diagram shown in Figure 1, it occurred to me that rather than randomly grouping scale items together to yield more continuous variables which improve SEM model fit, what the authors actually did was to use pre-established subscales for depression and the EFA-based two subscales for PIS. If that is in fact the case, I strongly recommend that the authors drop the use of the item parceling terminology and instead state in the measure descriptions that for purposes of this analysis they will be using the four distinct, pre-determined, literature-recommended subscales of depression as indicators of a latent depression factor and the two distinct subscales of PIS as indicators of a latent PIS subscale factor. In other words, these constructs aren't so much item parcels (which are often randomly grouped together item sets measuring the same construct); rather, these are pre-determined subscales, so the simplest solution here is not to describe these subscales as item parcels and instead describe them as subscales.

13. Data analysis, SEM goodness-of-fit: Because CFI and TLI come from the same family of baseline model comparisons and these two statistics are very close to one another in both construction and results yielded (CFI is pretty much a normed version of the TLI), I recommend dropping reporting TLI and just reporting CFI, RMSEA, and SRMR for
models containing all continuous variables and CFI, RMSEA, and WRMR for models containing at least one binary or ordinal endogenous variable (minor weakness).

14. Data analysis, SEM description: I was curious why the authors used the WLSM estimator for models with at least one binary or ordinal endogenous variable instead of the more well-studied default WLSMV estimator? (minor weakness)

15. Discussion, limitations: The authors correctly point out that results may be biased due to the exclusion of successful suicide attempters from their sample. However, they are silent on two additional limitations, which I think are also critically important to mention: 1) that those who refuse to participate might also be higher on stigma, depression, and potentially suicidality (and, relatedly, that those who are not visiting the clinic at all are also excluded - those who visit the clinic may be the most well-engaged and retained in HIV care) and 2) there is a growing body of statistical literature calling into question the utility of performing mediation analyses with cross-sectional data. Mitchell and Maxwell in their 2013 article "A Comparison of the Cross-Sectional and Sequential Designs when Assessing Longitudinal Mediation" in Multivariate Behavioral Research, 48(3), 301-339, provide a nice review of this literature. The degree of potential bias is strong enough to prompt some journals and journal editors to reject any paper submitted with a mediation analysis of cross-sectional data. I don't ascribe to such an extreme view - in the absence of longitudinal data, sometimes cross-sectional data, no matter how flawed it may be, is better than nothing and mediation analyses of cross-sectional datasets may provide an important first or preliminary look upon which subsequent longitudinal inquiries can be built. But I do feel it is crucial for authors of papers involving cross-sectional mediation results to acknowledge the potential bias inherent in such analyses in their limitations sections and would urge the authors of this paper to do the same and to go beyond the single boilerplate sentence stating that causal inferences cannot be drawn from a cross-sectional study. I think that's fine for a standard regression analysis, but when mediation is hypothesized, I contend the bar is higher for what conclusions can vs. cannot be drawn from such an analysis. (major weakness) This is yet another reason I'd recommend the authors subdivide their planned data analyses and then the results into more distinct descriptive and inferential subsections and spend more time discussing the important descriptive results. Doing so would enable them to recast the SEM as more exploratory given the cross-sectional nature of their data.

16. Table 1: Some of the items appear to have been cut off. Please edit the table to include the full item content (minor weakness).

17. Table 2: The sample size(s) should be listed, either in parentheses in the title or in a table footnote (if the same N is used for all results) or row-wise by variable if different sample sizes are used for the various variables. (minor weakness)

18. Tables 3 and 4 and 5: The sample size(s) for the correlations and the SEM should be listed. (minor weakness)
19. Tables 4 and 5, footnotes: The SE should be labeled as the "standard error", not "standardized error." (minor weakness)

20. Table 5: Please add rows listing the results for the measurement model coefficients (the factor loadings) (minor weakness).

21. Figure 1: By saying "All factors are significant", I assume the authors mean to say that "All factor loadings are significant"? (minor weakness)

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

No

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Yes

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

No

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

**Quality of written English**
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal