Author's response to reviews

Title: Detecting depression among adolescents in Santiago, Chile: sex differences

Authors:

Ricardo Araya (riaraya.psych@gmail.com)
Jesus Montero-Marin (jmontero@unizar.es)
Sergio Barroilhet (sbarroilhet@gmail.com)
Rosemarie Fritsch (fritsch.rosemarie@gmail.com)
Alan Montgomery (alan.a.montgomery@bristol.ac.uk)

Version: 3 Date: 11 March 2013

Author's response to reviews:

We include the answers to the reviewers' questions.

We wish to thank the reviewers for the attention shown to our work and for their interesting contributions, which have allowed us to improve the quality of the manuscript. Additions and changes made to the manuscript are highlighted in yellow.

Yours sincerely,

1. However, I think the discussion could be more informative if it utilized the findings of existing research on gender differences in depression.

This paper is not specifically about gender differences in depression but about sex differences in questionnaire cut-off points when used among adolescents; there is virtually nothing in this respect. It would be impossible to cover in great depth the theme of gender differences in depression in this paper, as I am sure the reviewer would agree. However there is a full paragraph dedicated to sex differences in page 4 and another full and long paragraph at the beginning of page 10. We have added also another comment in the conclusions in this respect.

2. In addition, a discussion of cultural influences in the design of the study itself (for example, questionnaires vs. interviews, fixed alternative vs. open ended questions, self-reporting vs. observing behaviors) can further raise possibilities for future research in the areas of cross-cultural and gender differences in depression.

This is a fair point but our decisions in the design were much more grounded on practical issues such as the most efficient way of measuring emotional symptoms with large samples participating in a RCT. In view of this using either interviews or observations were unrealistic given the size of the trial. We have added a comment in the methods and conclusions but there are already several
comments in this respect throughout the paper.

3. The title of the manuscript refers to gender differences, but the abstract refers to sex differences. These two terms are not interchangeable and it is important in framing the manuscript for the authors to be clear about which they are examining and then to be consistent in use of term and meaning throughout the manuscript. Sex refers to the physical characteristics with which an individual is born and is a reflection of their genetic constitution. Gender is a personal affiliation and may or may not reflect the sex one was born. It is the extent to which one identifies as male or female and exists on a continuum. It appears that the authors are interested in sex differences, but because the means of determining this variable in the sample is not given I am not entirely sure I have interpreted their intention correctly.

This is a fair point and we have decided to change ‘gender’ for ‘sex’ all throughout.

4. More information about the composition of the sample is needed. The authors report that 53.6% were girls in 10th grade, but provide no other grade and sex information. Since the study is focusing on sex differences in depressive symptoms this is important information to leave out.

All of the students were attending 10th grade (we have added sentence to clarify this) and this is the gender distribution of the total sample. We also provide mean age of the sample. We are unsure what more information the reviewer would like to include.

5. Can the authors provide an explanation and rationale for why the particular methods of determining which students in the two samples would participate in follow-up clinical interviews were chosen?

We have already done so in page 4.

The study sample consisted of 592 participants with a mean age 15.5 (SD=0.98), almost half (53.6%) were girls, all of them attending Grade 10th (approximately 10 years of education) in these schools. Two samples were drawn using different methods. The first sample of 250 students was drawn based on their BDI-II scores collected as part of the baseline assessment in five schools in the active arm of the trial. The first 50 students with BDI-II scores between 0 and 6 (lower tertile), the first 100 students whose scores in the middle tertile (7/15), and the first 100 students with high scores (>15) were invited for a clinical interview. For the second sample, all the 352 students in the control arm of the trial who scored high (#15 for girls and #10 for boys) on the BDI-II were invited for clinical interviews. Students answered the BDI-II in the classroom and clinical interviews were performed within 72 hours in a private office in the school for both samples. A trained clinician blinded to the student’s BDI-II status administered this psychiatric interview. In order to improve the blinding of the assessors, interviewers were rotated between schools, so that no-one who participated in the administration of the BDI-II in a particular school also interviewed in the same
school.

The reason to use this methodology for sampling, which is often used in validation studies, is to ensure there will be sufficient number of cases to contrast against the gold standard given our interest in having accurate estimates of sensitivity (the questionnaire is able to accurately identify as many as possible true cases).

6. At the top of the 4th page, 4th line down, the authors report rates of depressive symptoms in Chilean adolescents stating that "youngsters scored score 19". The second "score" should be deleted.

Done. Thanks.

7. In the d2. In the description of the BDI-II the authors initially correctly identify it as measure of depressive symptoms, but in reporting typical cut-off scores then refer to severity of depression. Because the BDI-II is not a diagnostic instrument, care should be taken to always refer to it as a symptom measure

The reviewer is correct that strictly speaking we should not call it severity of depression. However this is one of the aims of a validation study to establish the best cut-off points for a certain diagnosis and if possible to suggest the cut-off points that will match a range of severity of that condition. Nonetheless we have made some changes in the writing to emphasize this point.

8. Is there precedence in the literature for using only selected subscales of the RCADS in analyses? If so, please provide citations to support this decision. The theoretical rationale offered by the authors for this approach is fine, but altering the way in which a standardized scale is used can affect the psychometric properties of it.

No we were unable to find any other studies using a similar approach but several papers do report scores for the subscales. Nonetheless anxiety or RCADS were not the main aim of this paper but we just wanted to include it given the high co-morbidity of anxiety and depression and in an attempt to contrast the main findings concerning BDI-II.

9. For readers less familiar with psychometric analyses it would be helpful in the data analysis section to not only list the goodness of fit indices used, but to also provide the values considered acceptable on each.

We have added further information in the manuscript to answer this.

10. How is it possible for the mean BDI-II inter-item correlation to be larger than the mean for either of the sex groups?

We have revised these analyses and we can confirm that the results are correct. The most likely explanation is that inter-item correlations for males, females, and the total sample come from different polichoryc correlation matrices, which are included in the annexes.
11. Can the authors clarify their cut-off points? All are listed as a two-point option (e.g., 13/14) rather than the more traditional single score which is interpreted to mean that scores at that point or higher are considered clinically significant.

That’s exactly what we mean. The cut-off is established at 13/14, which means that values 14 and above represent cases and vice versa. We have added a clarification for this in the text and tables.

12. I wonder about the appropriateness of references not published in English for a manuscript submitted to appear in an English language journal (e.g., 36 and 37). Since the reason we cite others’ work is to allow readers to verify the claims we make in our manuscripts and to allow them to delve deeper into areas of particular interest it might be best to limit references to articles published in English. However, I recognize that the most appropriate references to use in some cases (e.g., 36) may only be available in Spanish and therefore trust the authors to decide how to address this issue.

We think an international journal should be able to accept references in other languages. Otherwise there is a serious risk of publication bias because only those people who can write in English would get their work published and given that this is a study with applications to Spanish-speaking populations we think it is appropriate to leave them in but we would leave it to the editors to decide.

13. Number of cases with depression is reported at table 4. It should be interesting having this information explicitly in the text too, additionally reporting standard errors, and tests of differences between groups.

As suggested we have included all this information in the text (Pg10, 2nd para).

14. Given the focus of the work in differences in depression according to gender, it should be very interesting to analyze measurement invariance of the BDI-II between men and women, for instance through multi-group confirmatory factor analysis. If you cannot assume such invariance, the rest of comparisons do not make sense, so if sample size allows performing that type of analyses they could be very informative.

This is an interesting suggestion which we agree with. We performed and include in the revised manuscript the results of this analysis.

15. In table 3, instead of communalities (which can be easily calculated from factor loadings) you should provide standard errors.

As per suggestion we have included the standard errors in the table. However we kept the commonality coefficients for the EFA for the total sample because it contributes further to understanding our conclusions.

16. If they are available, it should be useful knowing other fit indices, such as chi-square (with DFs) or RMSEA (with CIs). Likewise, updated cut-off criteria for fit indices should be considered. Information regarding sample size adequacy
would be helpful, such as Hoelter’s index.

As requested, we have calculated and included these results [chi-square (df), GFI, AGFI, RMSEA (90% CI), SRMR, Hoelter (0.05)].

17. Chronbach’s alpha is not considered an adequate estimator of reliability. Raykov’s rho, for example, could be a good alternative. See Bentler (2009), Alpha dimension-free, and model based internal consistency reliability, Psychometrika, 74, 137-143, for alternatives.

As suggested per reviewer, we have examined the reliability of the scale using congeneric, tau-equivalent, and parallel models, in the total sample and the sample divided by sex. We chose the model that fitted better with the data, applying GLS method, and establishing comparisons between models from the least to the more restrictive, through ##2. The reliability value was estimated by squaring the implied correlation between the composite latent true variable and the composite observed variable, to arrive at the percentage of the total observed variance that were accounted for by the “true” variable. All results are presented in Table 5.

18. Clinical diagnosis was performed for only one person. This limitation should be explicitly included in the discussion.

This was a misunderstanding what we meant is that the person was a trained clinician but there were 3 of these trained clinicians. We have corrected this.