Reviewer’s report

Title: Diagnostic validity of the MINI-KID disorder classifications in specialized child and adolescent psychiatric outpatient clinics in Sweden

Version: 0 Date: 29 Oct 2018

Reviewer: Laura Duncan

Reviewer's report:

This manuscript aims to examine the diagnostic validity of the MINI-KID in general child and adolescent psychiatric care. I am not sure whether this can be achieved using the methods proposed. We do not have a 'gold standard' criterion in the measurement of child psychiatric disorders. A comparison with an initial diagnosis seems likely to result in poor agreement as standardized interviews were introduced in response to the poor reliability of clinical diagnoses, but the authors are not clear about whether they expect good agreement or not. Regarding the LEAD diagnosis after 6 months, this appears to be an evaluation of predictive validity and I don't know how useful this is. It is unclear how we can expect this to achieve agreement with the MINI-KID interview given the time lapse and the likelihood that symptoms/concerns will have attenuated/improved with treatment over that time. I am not very familiar with the LEAD diagnostic process and there are issues with how the process is described in the manuscript so I may be misunderstanding how this assessment was conducted. I am unclear about: 1) whether all or part of the patient data was used as the authors say "part of all data" in a couple of places; 2) whether the assessment is meant to better capture the initial/presenting diagnosis or if it is the diagnoses that best represents the 6 month period or the end of the 6 month period; and 3) how the discussion process was used and what happened if there was disagreement between experts. If the LEAD diagnosis is meant to serve as a reliable diagnostic assessment then I think more explanation and description is needed. In this case, an evaluation of convergent validity between the LEAD diagnosis and the MINI-KID may be possible. If the LEAD is to be considered a 'gold standard' then evidence is needed to support this and in that case an evaluation of diagnostic validity may be possible but I am unable to judge based on the information provided. Dropping the use of the initial clinical diagnosis and a better explication of the LEAD diagnoses as a criterion and focusing on a convergent validity analysis could make a better contribution.

I draw quite different conclusions from the results that the authors do. The low levels of agreement (kappa) speak to the unreliability of clinical diagnoses and also these types of standardized interviews. Our team showed this in a recent systematic review and meta-analysis of standardized interviews (Duncan, L., Comeau, J., Wang, L., Vitoroulis, I., Boyle, M. H., & Bennett, K. (2018). Research Review: Test-retest reliability of standardized diagnostic interviews to assess child and adolescent psychiatric disorders: a systematic review and meta-analysis. Journal of Child Psychology and Psychiatry.). However, arguments have been made that we can not expect better, (Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E., & Regier, D. A. (2012). DSM-5: how reliable is reliable enough?. American Journal of Psychiatry, 169(1), 13-15.) and I think the others can evaluate their findings in this context and give their evaluation criteria in their methods section.
The other major comments/concerns include:

1) It is not clear what the precise setting/purpose the MINI-KID is being evaluated for in this paper. The authors mention 'general child and adolescent psychiatric care' but also 'Child and Adolescent Psychiatry'. It would be helpful to provide more description and clarify whether they are referring to only clinical settings and what type of clinical settings (hospitals, community agencies, primary care) and by whom (only psychiatrists?).

2) I think that kappa is the appropriate measure of agreement in this analysis although I agree that reporting % agreement is also useful for evaluating the impact of low prevalence on kappa. However, low prevalence doesn't appear to be too much an issue for the majority of disorders. With regards to using Gwet's AC1, I don't think that correcting kappa in this way is useful. Low prevalence is a feature of mental health disorders and is a problem for trying to assess disorder with these types of interviews. Researchers/clinicians have to be aware that using these types of interviews will result in low levels of reliable variance. It is therefore misleading to correct the kappa estimates for low prevalence as it is unclear what the implications of this are for the usefulness of the instrument. So long as prevalence is reported in addition to kappa, the reader is aware of the issue of low prevalence and in most cases, there are a reasonable number of test positives.

3) It is unclear how information from parent and youth informants was combined in all three diagnoses. For the MINI-KID were the parent and youth jointly interviewed and scored separately or were they interviewed separately? It is unclear how the separate interviews were administered. Was this decision made on patient preference or was it based on some systematic process such as age of the youth? Given the low agreement between parent and youth diagnoses, this is clearly in important consideration when conducting these types of interviews.

4) There is not enough discussion about what criteria is being used to evaluate agreement. The authors mention reasonable agreement in the abstract and the discussion but don't quantify this. It would be helpful if the authors could outline how they are going to conduct the evaluation and what criteria they will be using in the methods.

5) The authors conclude that the MINI-KID is a cost-effective tool but the paper mentions nothing about cost. I understand the point they are making about low-burden (the tool is short and comprehensive for its length) but my understanding is that there is a per use cost associated with the tool so I don't know if they can comment on cost-effectiveness without going into this in the analysis.

6) The authors talk about the reliability and validity of the MINI-KID but I believe it is best practice to discuss reliability and validity as properties of the classifications produced by the instrument, not the instrument itself.

7) Can the authors add some specific research objectives/questions to the 'Aim'. What specific questions are they trying to answer?

8) The statistical methods section refers to inter-rater reliability but I don't think that is what is being assessed here. If it is, then these needs to be explained and described more explicitly.

9) Table 1 needs to be formatted as a sample characteristics table. It is currently very difficult to interpret.

10) I think there needs to be some discussion on the limits that low reliability places on validity as this can help explain the low levels of agreement.

11) The fact that the sample was not random is not a limitation here I don't think as the study objectives are to look at within individual agreement on the different assessments. I think representativeness would be important if the authors expect validity to differ based on some demographic characteristic. They are clear that this is a clinical sample. It is not clear if it was a convenience or consecutive sample from the methods however, perhaps they can specify this.
12) Tables 5a and 5b are hard to interpret without specific prevalence estimates and the +/+, +/-, -/+ , +/- counts. Unless these are specific objectives for the paper, I would drop these tables. If they are retained, I would format them the same way as Table 4. Also, 'younger' and 'older' needs to be replaced with specific age ranges.

A few minor comments that can easily be addressed include:

1) There are differences between the use of MINI-KID, MINI KID, MINI-KID interview, the MINI-KID. It would be helpful to use the same format of acronym for consistency.
2) There are typos and grammatical issues throughout. The manuscript would benefit from more proof reading and correction.
3) The background discusses the impact of missing information on diagnostic accuracy but it is unclear how the MINI-KID corrects for this. Respondents do not have to answer all the questions in the MINI-KID. Standardized interviews were introduced to address the unreliability of clinical diagnoses and I think this is worth mentioning as this provides a rationale for using the MINI-KID.
4) 'takes much less time to administer'-can the authors provide some details on the time reported in the literature to date and compare it with other interviews?
5) The authors mention reliability in a few places but it is not clear what kind of reliability this is (inter-rater, test-retest). It would also be helpful to be specific about estimates of reliability and validity that have been reported in the previous literature (e.g. page 4, line 7).
6) Can the authors better explain what a 'regular initial clinical diagnosis' is and how it is formulated?
7) In the methods it would be helpful for the authors to describe what their expectations/hypotheses are about how the various assessments will differ in terms of validity and reliability and how they will measure up to each other.
8) Is the sample size appropriate for the study objectives?
9) Page 5, line 6 'the unit' - what is this referring to?
10) Page 5, line 7,8- this sentence is unclear. Also using 'regular' is hard to interpret when the reader doesn't know the context.
11) How many interviewers were there at each site and how was blinding handled?
12) 'It is supposed to be sufficient as a short yet reliable interview that can be used both for research and clinical purposes.' This sentence would benefit from a reference.
13) The authors explain how the MINI-KID can be used with different respondents but don't explain the procedure that was used for individual vs joint interviewing in this study. It would also be helpful to briefly describe how it was administered and scored (paper/pencil, scored in real time or after the interview etc.)
14) Page 6, 'Shehaan' should be 'Sheehan'.
15) Page 8, line 18 onwards: I am unclear about the importance/relevance of the discussion about visit numbers and this is not included in the analysis section. Does this impact the LEAD analysis? More details about the importance of this information is needed for the reader.
16) I would exclude the diagnoses not assessed by the MINI-KID as there isn't a clear reason to include them. Otherwise, this reason should be provided.
17) Can the authors list the specific MINI-KID disorder modules and how they were grouped? I believe the anxiety disorders are in different modules, which anxiety disorders were included?
18) Were there any issues with missing data in the MINI-KID? If so, how was this handled?
19) The tables also refer to interrater agreement, I don't think that is what is being assessed.
20) There are some items in the results that are not included in the study aims or analysis (e.g. visits, agreement by age group, parent-youth agreement). These should be outlined as specific study objectives and a rationale provided for why they are important or they should be excluded
or included as supplementary information if the authors think the information is relevant/useful/important.

21) Why would the authors expect agreement to be different by age group if the tool has been validated across the age span?
22) Were the respondents reimbursed for their time? If so, can that be added to the Acknowledgements?

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