Author’s response to reviews

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

Authors:

Camilla Högberg (camilla.hogberg@sll.se)
Eva Billstedt (eva.billstedt@gnc.gu.se)
Caroline Björck (caroline.bjorck@sll.se)
Per-Olof Björck (pobjorck@hotmail.com)
Stephan Ehlers (stephan.ehlers@vgregion.se)
Lars-Henry Gustle (lars.henry.gustle@skane.se)
Clara Hellner (clara.hellner@ki.se)
Henrik Höök (henrik.hook@sll.se)
Eva Serlachius (eva.serlachius@ki.se)
Mats Svensson (mats.a.svensson@skane.se)
Jan-Olov Larsson (jan-olov.larsson@ki.se)

Version: 1 Date: 28 Jan 2019

Author’s response to reviews:

Answers to the Editor:

Editor Comments:

I was fortunate to receive comments on your paper from two experts in the field. I also read it carefully myself. As you will see the two expert reviewers’ reactions were mixed. Based on these reactions, I am inviting you to resubmit the manuscript, but I want to emphasize the importance of addressing all the reviewers’ critiques. I will summarize some points I consider important presently.

1. Sampling strategy and sample characteristics should be detailed to a greater extent; similar considerations held for the study procedure.
Sampling strategy, sample characteristics and study procedure is now described in more detail: page 5, line 17-24, page 6, line 1-10, 15-28

2. The multicentric nature of the study was not considered in the present manuscript. Please, take into consideration the hierarchical structure of the data (subjects are nested within sites) and/or explicitly acknowledge this limitation of your study.

The lack of multilevel analyses is now explicitly acknowledged as a limitation of the study: page 16, line 3-6

3. Please, provide a stronger rationale for relying on the LEAD approach. I have to admit that I agree with Reviewer 1 that a comparison with “initial clinical assessment and preliminary / tentative diagnosis” is likely to result in poor agreement with a standardized interview. Please, provide clear hypotheses for your study and presents results according to them.

In the introduction we have revised the text to in order to provide support the usefulness of LEAD and the rationale for using this method page: 4, line 23-29. The aim of the study is revised: page 5 line 6-13

4. Please, temper your language regarding the conclusions of the study.

The conclusions have been revised: page 16 line 16-26

5. The writing needs to be improved and the manuscript should be reviewed for typos and language and grammar idiosyncrasies. Please, avoid sentences like, "The impression was that participating...".

The manuscript has been reviewed for typos and language by Anna Spyrous from UK.

BMC Psychiatry operates a policy of open peer review, which means that you will be able to see the names of the reviewers who provided the reports via the online peer review system. We encourage you to also view the reports there, via the action links on the left-hand side of the page, to see the names of the reviewers.

Answer to reviewer1

Reviewer reports:
Laura Duncan (Reviewer 1): 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.

We used research diagnoses according to the LEAD procedure not as a 'gold standard' but as a method to come closer to 'true' diagnoses. Spitzer suggests the use of multiple sources of information and monitoring of the patient’s condition and diagnosis over time. In the absence of an established gold standard, LEAD has been widely used in psychopathology research for studying validity of diagnostic procedures The introduction is revised in order to underline that several studies, added to the reference list, have used LEAD in this way, evaluating different instruments and categories of diagnoses: page 4, line 23-29
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.

We agree with this remark and have changed the wordings in the Aims section. We still think that it could be of interest to present this comparison between the initial clinical diagnoses and MINI-KID.

The aim of the study is revised: page 5 line 6-13

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.

We agree that 6 months is a long time for these patients. But, if you have a method with good predictive validity that method could be positive and useful.

Actually we had a shorter evaluation period of 3 months for the LEAD assessment in the first version of the study protocol but changed to 6 months. The reason was that for most patients it often did not happen much between 6 weeks (the initial clinical diagnoses) and 3 months as reflected in the documentation in the charts. One of the advantages with LEAD is monitoring of the patient's condition and diagnosis over time and retrospectively assess the diagnoses.

Another aspect of time is the setting. The outpatient clinics are at the secondary level of care (now described in the methods section). Many patients had problems for a long time, for example developmental disorders with often a rather stable symptoms and reduction functional level.

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;

The LEAD assessment included all data. The wording (took part) has been revised: page 8 line 18.

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

The LEAD assessment is meant to better capture the diagnoses that best represents the 6 month period. This information is now added (page 8 line 25).

3) how the discussion process was used and what happened if there was disagreement between experts.

The paragraph about the LEAD procedure is now revised: page 8 line 23-30

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.

More information about the LEAD procedure is added in the Introduction and the Methods section. We hope this will give enough support for the evaluation of convergent validity between the LEAD diagnosis and the MINI-KID.

We still think that it could be of interest for the potential readers to present this comparison between the initial clinical diagnoses and MINI-KID. We agree with the reviewer that this is not a validation of MINI-KID and revised the wordings in the Aim section.

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.).

Thank you for the information about this new comprehensive and important review! It is added to the introduction and the discussion. The conclusions have been revised:

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 authors can evaluate their findings in this context and give their evaluation criteria in their methods section.

This helpful information has been added to the methods section and we now try to interpret our findings according to their evaluation criteria:

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?).

Information about the clinical setting is now described in more detail in the first part of the Methods section:

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.

The wordings about kappa and Gwet’s AC1 have been revised in the methods section in order to correspond more with the views of reviewer 1. Since this is an area of discussion we think it would be valuable if the Gwet’s AC1 still could remain in the Tables. We will change this if the editor and reviewers want to completely omit Gwet’s AC1 from the manuscript: page 10 line 2-11.

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.

“The interview often takes place with both parent and child present at the same time, but it can also be done with the young person separately”. Separate interviews with the informants were not performed. Page 6 row. The MINI-KID procedure is described in more detail on page 7 row 7-17 and page 8 line 1-11. The versions used and the participants in Table 1.

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.

We now outline how to conduct the evaluation and the criteria in the end of the Methods section: page 9 line 26-28 and page10 line 3-11

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.

The costs are mainly related to the working-hours for the staff. But, we agree, the costs are not analyzed and presented enough. The wordings about cost are removed from the revised manuscript.

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.

We agree, the wordings are changed to “MINI-KID disorder classifications” throughout the text including the Title.

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

The Aim section is now revised and includes more specific research objectives/questions: page 5 line 6-13

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.

The statistical methods section is now revised and “inter-rater reliability” is removed from the text.

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

Table 1 is removed and the figures are presented in the first part of the Results section: page 11 line 2-5

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.

Now there is a paragraph in the Discussion section bout this important issue: page 15 line 8-17

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.

The Limitations section is now revised accordingly and the sentence about “not random” is removed.

The sampling procedure leading to a convenience sample is now described: page 6 line 1-10 and also in the Limitations section: page 16 line 6-8.

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.

Table 5a is removed. Since it was an aim of the study to compare the two versions of MINI-KID table 5b (now Table 4) is retained in the revised manuscript. We hope there is enough information using the present format in Table 4 for comparing the measures of agreement between LEAD and the two versions of MINI-KID, MINI-KID-S (49 patients) and MINI-KID-P (51 patients).

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.

In the text we now use MINI-KID as the general designation for the SDI, MINI-KID-S and MINI-KID-P when specifically describing the standard and parent version: page 7 line 27-28

2) There are typos and grammatical issues throughout. The manuscript would benefit from more proof
reading and correction.

We have revised the manuscript including more proof reading. The manuscript has been reviewed for typos and language by Anna Spyrous from UK.

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.

This is now discussed more in the introduction: page 3 line 5-6 and 16-17. MINI-KID uses two to four screening questions for each disorder. If the screening questions are positively answered, additional symptom questions are given for the particular disorder. This is in contrast with an unstructured clinical interview having the risk to terminate the interview before exploring all alternatives. (According to clinical experience if you have a reasonable good contact with the patient and parents, most often the respondents answer the questions in MINI-KID (and also in unstructured clinical interviews)).

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?

Such information is now included in the Introduction (page 3 line 22-24).

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).

We have now added information about type of reliability: page 4 line 9-12 nd line 17-21
Information about estimates has been added on page 9 line 26-28.

6) Can the authors better explain what a 'regular initial clinical diagnosis' is and how it is formulated?

The setting and 'regular initial clinical diagnosis' are now described in more detail in the first part of the Methods section: page 5 line 17-24

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.

Such expectations are now added to the statistical methods section.

8) Is the sample size appropriate for the study objectives?

It should be appropriate for the most prevalent diagnoses.

9) Page 5, line 6 'the unit' - what is this referring to?

Unit was changed to “clinic”: page 5 line 31
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.

The text about the procedure and the context is revised: page 5 line 17-24

11) How many interviewers were there at each site and how was blinding handled?

The number of interviewers in the three sites were: Stockholm n=11, Gothenburg n=4, and Skåne n=2. This information is added to page 6 line 31

The clinician at the unit performing the clinical assessment was "blinded" for the outcome of the MINI-KID interview and did not receive any information about the results of the interview. The outcome of the MINI-KID interview was not returned directly to the patient/parents and not included in the medical record. However, the patient/parents was offered to be informed of the results of the MINI-KID interview after the LEAD assessment was completed: page 6 line 22-28.

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.

A reference is inserted: page 7 line 21.

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.)

This is described on page 8 line 1-11.

14) Page 6, 'Shehaan' should be 'Sheehan'.

Sorry for this, now changed.

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.

The sentences about visit numbers are moved to the Methods section describing the LEAD procedure. The idea is the give more information to the reader what information the LEAD assessment is based on: page 8 line 24-30.

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.
We think it could be of interest for the reader to include them. These sentences have been added to the Methods section about LEAD: “The initial clinical assessment and LEAD assessments resulted in some instances to diagnoses not covered by MINI-KID. This is described in Table 1.”

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?

The MINI-KID disorder modules are now listed in the Methods section: page 7 line 8-18. All the anxiety disorders were included and summarized in the Tables as “Any anxiety disorder”.

18) Were there any issues with missing data in the MINI-KID? If so, how was this handled?

The interviewers were given the freedom to stop posing questions when they thought they had enough information (although the instruction according to MINI-KID is to ask all follow-up questions for scoring if screening is positive). After an assessment of the answers, all the interviewers completed the form with diagnoses on the first page in the MINI-KID questionnaire. However, each of the single modules were not always completed by the interviewers. So missing could then mean “asked but not written in the module” or “not asked for at all”. The analyses are based on the form with diagnoses on the first pages (page 2 and 3) in the MINI-KID questionnaire.

19) The tables also refer to interrater agreement, I don't think that is what is being assessed.

This is changed to “Inter-agreement between MINI-KID and the initial assessment” in Table 2 and “Inter-agreement between MINI-KID and LEAD” in Table 3.

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.

“Visits” see 15 above.
“Agreement by age group” this table is omitted from the revised manuscript see “The other major comments/concerns” see12
“Parent-youth agreement” The comparisons between MINI-KID-S and MINI-KID-P are outlined as specific study objectives: page 5 line 11-13.

21) Why would the authors expect agreement to be different by age group if the tool has been validated across the age span?

“Agreement by age group”is removed.

22) Were the respondents reimbursed for their time? If so, can that be added to the Acknowledgements?

The parents were not reimbursed for their time but the children were given movie tickets. This information is added to methods section: page 6 line line 27-27.
Saundra Stock (Reviewer 2): I feel a description of the "LEAD" diagnostic model earlier in the paper would be beneficial for those who are not familiar with it.

A more detailed description of LEAD (and more references) is now in the Introduction page 4 line 23-29 and in the Methods section page 8. In the introduction we have revised the text to in order to provide support the usefulness of LEAD and the rationale for using this method page: 8 line 13-30.

Specific Recommendations:
Page 3, line 3: Change to "Clinical interviews in Child and Adolescent Psychiatry obtain information about symptoms and impairment related to various disorders, however, missing diagnostic information often gives..."

This has been changed.

page 3, line 30: should there be a "," after et al.

This has been changed.

page 4, line 4: insert a recent study of "the" MINI-KID

Ok! Done.

page 4, line 5: DSM-IV-TR disorders (add the "s"). Very long sentence that is hard to follow. recommend ending after "general population samples". It compared the factor structure...

Ok! It is now revised, thank you.

Page 5, line 19. "were examined by three MINI-KID interviews". Does this mean that each person conducted 3 interviews to be certified by the MIN-KID trainers or that 3 trainers assessed the interviewers?

Each person conducted 3 interviews to be certified by the MIN-KID trainers. This is now revised.

Page 6, line 17: "took part of all data" - can you reword? Does this mean two clinical experts reviewed all data in the patient's medical record?

Two clinical experts reviewed all data in the patient's medical record. This is now revised page 8 line 18-19.

Page 8, line 4: place inside ( ) (median 1 SD 1.3, min 1, max 8).

Ok! Done.

Page 8, line 8:average 4 interviews each. Is this per patient? Suggest "average 4 interviews per patient"
The sentence is changed to: “Each interviewer (n=16) performed on average MINI-KID-interviews with four patients.” The sentences about visit numbers are moved to the Methods section describing the LEAD procedure. The idea is the give more information to the reader what information the LEAD assessment is based on: page 8 line 24-30.

Page 8, line 8: insert "," following During the interview,

Ok! Done.

Page 8, line 9: change to "At least one parent participated in the interview for 88 of 101 patients".

The text is changed to: In 88 of 101 interviews at least one parent participated together with the patient.

Page 8, Line 18: Change to During that period "," the total number of visits after the first visit "was" on average 7.5 (median 7, SD 4.3, min 2, max 21 with the 25th percentile at 4 visits).

Ok! This is changed. The paragraph is now moved to the Methods section describing LEAD.

Page 8, line 22: change "on" to "receiving"

Ok! This is changed.

Page 8, line 28. Change to "The MIN-KID interview yielded a higher number of diagnoses especially for anxiety disorders, behavioral/disruptive......"

Ok! This is changed.

Page 10, line 2: change "gives" to "provides"

Ok! Done.

Page 11, line 18. "problems not may be asked for" - makes no sense. Reword. Perhaps "that may not be asked about or documented"

Ok! This is changed.

Page 12, line 2: Our result or Our result(s)?

Changed to Our results

Tables look fine.

Thank you!
We thank editor Antonella Somma and the reviewers Laura Duncan and Saundra Stock for relevant questions and helpful remarks.