Author's response to reviews

Title: Predictive Properties of the A-TAC Inventory When Screening for Childhood-onset Neurodevelopmental Problems in a Population-based Sample

Authors:

Tomas Larson (tomas.larson@med.lu.se)
Sebastian Lundström (sebastian.lundstrom@neuro.gu.se)
Thomas Nilsson (thomas.nilsson@neuro.gu.se)
Eva Norén Selinus (eva.noren-selinus@sll.se)
Maria Råstam (maria.rastam@med.lu.se)
Paul Lichtenstein (paul.lichtenstein@ki.se)
Clara Hellner Gumpert (clara.gumpert@ki.se)
Henrik Anckarsäter (henrik.anckarsater@neuro.gu.se)
Nóra Kerekes (nora.kerekes@neuro.gu.se)

Version: 3 Date: 26 April 2013

Author's response to reviews: see over
Dear Editors,

We would like to express our gratitude for your constructive critics and helpful input for the improvement of this manuscript, which has now undergone a point-by-point major revision in response to the feedback from reviewers and editors. Throughout this cover letter the responses to specific comments are in italics, and crucial revisions in the text are highlighted in grey. Please note that the references to pages and lines in the reviews are according to previous submission and may possibly not coincide with the pages and lines of the re-submission.

Yours sincerely,

Tomas Larson

EDITORS REQUEST:

Please re-word your consent statement on page 8 to clarify that all children and parents in CATSS gave informed consent.

This has now been clearly stated as requested (page 8):

The overall response rate was 80%, and all subjects were protected by the informed consent process, and repeatedly given the possibility to withdraw their consent and discontinue their participation.

Referee number 1.

Minor Essential Revisions

• P18 line 10: “…. the sensitivity and specificity of screening tools are not affected by the prevalence of the disorders…” is right. But in the previous paragraph, the authors wrote “A diagnosis for a rare condition in the general population………may theoretically result in high sensitivity and specificity…….” Are “sensitivity and specificity” mistakes?

They were, and this has now been corrected (page 18).

• The first sentence in discussion;

“A-TAC screen-positive children almost uniformly had an NDP…..” From the Table1, the number of A-TAC screen-positive children is 247, and out of them, 125 had no NDP, that is 222 had NDP. It does not appear to be “almost”.
About half of all A-TAC screen-positive children had an NDP at clinical examination, and about 40% of them turned out with the same NDP diagnosis three years later.

- Discretionary Revisions

The information of administration time and potential interviewers (occupation and training etc) are very important for a screening tool. So, those had better be described.

The screening was conducted over the telephone by a market research center, “Intervjubolaget,” using a computerized version of the A-TAC inventory. All interviews were performed by laypersons, with just a brief introduction to research methodology and child mental health problems. The interviewers entered the responses directly into an electronic database.

The average interview time for screening the general population in the CATSS with A-TAC is 27.5 minutes.

- Table 1;

32/157 screen-negative siblings had other K-SADS Diagnosis (no NDP), and 13/46 screen-negative random controls had it. What symptoms are they?

We realize that this is not very clear as stated, but “Other mental health problems” is defined as OCD and/or ODD and/or CD and/or ED, with no NDP overlap, and “Other K-SADS diagnosis” is defined as “Other mental health problems” PLUS and/or depression, anxiety, stress disorder, mania, and/or psychosis. This is in the footnotes to Table 1., but has now been reworded in hope of making it clearer:

Other mental health problems defined as OCD and/or ODD and/or CD and/or ED, with no NDP overlap
Other K-SADS diagnoses (i.e. “Other mental health problems” and/or depression, anxiety, stress disorder, mania, and/or psychosis with no NDP overlap)

Referee number 2:

Major compulsory revisions:

I agree very much with the authors that the NPDs are overlapping in nature and the ESSENCE model that indicates that the main “face” of the disorder may shift with time, development and psychosocial environment for an individual. As the design has a three year lag, considerable changes may thus be expected in which exact disorder symptom
criteria are met between those times. That this is a major effect in the study, and leading to an underestimation of the screening properties of the A-TAC is further supported by the relatively stronger precision of the instrument vs. ASD, which is also one of the most constant NPDs, at least between ages 12-15. Therefore I would suggest that

1. the paper is rewritten, putting more prominently this ESSENCE model and development (introducing it already in the introduction, not only in the discussion). The introduction and the paper need to stress that we cannot expect very high levels of specific diagnostic stability over three years and that the results will be accordingly. What we can expect, or what is to be expected, according to the ESSENCE model, is that the youths may still have another main NPD three years later, which, from the both the clinical and screening point of view, is more important.

This has now been addressed in the introduction (pages 4-5):

The course of the NPDs is not stagnant, and it has been suggested that the NPDs are not discrete entities at all, and that “it would be inappropriate to diagnose one problem and not consider the implication of the other(s)” [2]. Young children with problems severe enough to warrant clinical examination can be said to suffer from an “early symptomatic syndrome eliciting neuropsychiatric clinical examination” (ESSENCE), which may later in life meet criteria for an NPD or any mixture of NPD diagnoses. The ESSENCE model recognizes that children with NPDs are at risk of developing various functional impairments, mental health problems, and/or other difficulties, often requiring life-long interventions from medical and social services.

2. Therefore, I suggest that a further analysis is included; sens-sp, ROC and DOR for screening positive for any NPD vs being diagnosed with any NPD three years later.

This is a very good idea, but it turned out to be difficult to accomplish within the framework of this manuscript, (we have tried and failed, this is why we felt obliged to ask for more time). Since all diagnostic domains in A-TAC have different cut-off values, these must be converted into a uniform scale with a comparable agreement so as to calculate a theoretical composite measure, in order to make it possible to present sensitivity, specificity, NPV, PPV, and DOR for the entire NPD-group. Nevertheless, this is highly relevant, and should perhaps be addressed in a paper of its own, using A-TAC as an “ESSENCE-screening tool” of sorts. However, we calculated the percentage of screening positive for any NPD vs being diagnosed with an NPD three years later and added this to the results (page 13):

Out of all the 198 children who screened positive for any NPD diagnosis, 108 (55%) received at least one NPD diagnosis at follow-up.

3. I also lack some mentioning of the general rate of agreement that one could expect with three years of lag, and adding to that the usual inter-
rater agreement problems, in the introduction of the paper (and of course in the discussion again.)

The limitation has now been touched upon in the introduction as well (page 5.) This, of course, is a very important point, so much so, that we have devoted an entire article, recently submitted, solely to test-retest reliability of A-TAC.

There is a need for longitudinal follow-up studies of representative samples with NDPs that do not suffer from just assessing one diagnostic category.

4. I also find the long explanations of what ROC, DOR, accuracy etc are a bit disturbing. I know that although PPV etc are supposed to be entities that people understand, many don’t and that some introduction may be required, but in my opinion this takes too much space and distracts from the main focus. The authors could perhaps be more assertive and brief, and please add some references to central papers and literature on this for the interested reader.

This section has been shortened, and hopefully the remaining information now comes across as more self-assured (page 12).

5. I also suggest that accuracy is omitted altogether, as DOR is a more relevant measure in this setting, being prevalence independent.

The measure “Accuracy” has been omitted from the Table 2. and in the sections of Methods and Results (pages 12, 14-15).

6. When it comes to the rationale for the A-TAC, it comes after a rather long mentioning of other diagnostic instruments and their properties. I assume the authors are trying to say that the A-TAC fills an important gap, and I wish they would be more concise and clear about that. There is no need for describing the K-SADS, DISCO and ADI-R/ADOS at such length, but rather (again) be more assertive and clear about what the message is!

The section describing other diagnostic/screening tools has been shortened, and more assertive and clarifying sentences have been added (page 6):

There is, nevertheless, a continued need for an easily administered screening tool that covers the whole field of developmental problems and that can identify the broader spectrum of NDPs, revealing not only “caseness” but also subthreshold traits and overlapping conditions.

7. There is also some repetition in describing the A-TAC in the introduction and the methods section, please reduce this so that the same information is not repeated. The more general stuff can be moved entirely to the
introduction, and the specifics remain in the methods section.

*The description of A-TAC has been pruned (page 9), and some sections moved to the introduction (pages 7-8).*

8. When it comes to diagnosing using the K-SADS if I have understood it correctly, the authors have not used it at all as intended. The authors claim that they used K-SADS diagnosis as “From the clinical evaluation a K-SADS-PL diagnosis were attributed if the adolescent had 3 or more on a 5 point scale, 3 representing threshold criteria”. Surely, this is not how the K-SADS should be used. And I wonder what is meant by 5-point scale. For which disorder? For all disorders? I would like some more explanation and rationale for this way of using K-SADS.

*This was unclear in the original manuscript, and we thank the reviewer for pointing this out. The K-SADS was used according to its directive, but this was poorly articulated. The instrument contain a “Summary Lifetime Diagnoses Checklist”, were each diagnosis is finalized and attributed a numerical code: 0-4, indicating: 0=No information, 1=Not present, 2=Probable, 3=Partial Remission, and 4=Definite, thus not a 5 point scale as such. In order not to confuse the matter further, we suggest that this section is omitted entirely as well, now we just point out that the K-SADS algorithm was used in order to establish a final diagnosis, with no further elaboration. This is now succinctly stated on page 10:*

Based on all sources of information available, the interviewer subsequently made a diagnostic summary rating according to the DSM-IV-TR criteria and the K-SADS-PL algorithms.

9. The final diagnosis was made using all available information by “a senior clinical expert”. Who is that, and was it the same or a few for the entire sample?

*The clinical expert was one of the coauthors, Dr. Eva Norén Selinus (ENS); she reviewed and finalized all diagnoses of the entire sample, together with each of the investigating psychologists. This is now stated in the manuscript (page 10):*

As a final step, the clinician, together with a senior clinical expert (ENS, a clinical board-licensed specialist in child and adolescent psychiatry), scrutinized the summary ratings for all cases assessed by the clinicians, with the aim of ascertaining definitive diagnoses based on and covering all accessible information from the clinical evaluations.

10. In the discussion: The authors stress the lower caseness screening properties for ADHD, LD and TD, but I disagree with this. It is surprisingly high, considering the methodology and natural development of children, and I think it gives the wrong impression to frame it as otherwise.
We have now presented these findings a tad bolder (page 15):

However, as expected in a nationwide study population, the instrument showed somewhat poorer screening test performance when screening for “caseness” in other than ASD diagnoses (ADHD, LD, and TD). Still, the “Low” cut-off values showed a sensitivity around 70% (TD here represents the low point with 45%), although without drastically improving specificity (64%–93%). Considering the interval of three years these figures are surprisingly consistent, and conform to notions that about half the children diagnosed with for example ADHD, seem to grow out of it [31], or that key ADHD features transform into other mental health problems [4, 32].

11. The low PPV is also highlighted in the discussion, but it isn’t low at all. It is quite high for this field of research, and perhaps it should be omitted completely, as it is dependent on prevalence.

We agree and therefore rephrased this part of the discussion (page 16):

A high rate of false positives is not uncommon in behavioral screening in which low positive predictive values are often reported [33].

Minor essential revisions/ discretionary revisions:

1. The tables are very clear and informative. But I would like to see the ROCs rather than just the AUC numbers, as the graphs very informative visually (and also makes it possible to see how the changes in cut-off point changes the properties of the instrument.)

ROC graphs have been added to the manuscript (see Figure 2-5). We will leave it to the Editors discretion to decide weather all should be used, or just ASDs and AD/HD.

2. It says the professional interviewers were given a brief introduction, I would like a specification of that.

See the previous response to Referee number 1 above.

3. In the limitations section, I think the rather low numbers of controls should be mentioned, and how that affects sens, spec, PPV and NPV.

The text now reads (page ):
4. Please spell out P.A.R.I.S.

It is now completed on page 11:

…the Paris Autism Research International Sib pair study (P.A.R.I.S.) pro forma.

5. Please include reference to the AUC values evaluation (poor-fair etc)

The AUC-evaluations are described in Statistical Analyses (page 11.), and now also declared in Results.

If you choose to maintain Accuracy as measure I think the definition must be corrected (top page 13)

Duly noted, however, we have chosen not to present “Accuracy” following the major revision advice point 5, above.

6. The first section of the discussion I think should be moved to the introduction, and the first section of the discussion should present the findings specifically as related to the main aims.

The first section of the discussion have been rephrased according to the points made by Referee number 1., we hope this will also satisfy Referee number 2.