Author's response to reviews

Title: The impact of study design and diagnostic approach in a large multi-centre ADHD study: Part 2: Dimensional measures of psychopathology and intelligence.

Authors:

Ueli C Müller (u.c.mueller@bluewin.ch)
Philip Asherson (Philip.Asherson@iop.kcl.ac.uk)
Tobias Banaschewski (tobias.banaschewski@zi-mannheim.de)
Jan Buitelaar (J.Buitelaar@psy.umcn.nl)
Richard Epstein (Ebstein@netvision.net.il)
Jaques Eisenberg (Jaques@cc.huji.ac.il)
Michael Gill (mgill@tcd.ie)
Iris Manor (Imanor@clalit.org.il)
Ana Miranda (Ana.Miranda@uv.es)
Robert D Oades (robert.oades@uni-due.de)
Herbert Roeyers (Herbert.Roeyers@UGent.be)
Aribert Rothenberger (arothen@gwdg.de)
Joseph Sergeant (JA.Sergeant@psy.vu.nl)
Edmund Sonuga-Barke (ejb3@soton.ac.uk)
Margaret Thompson (M.Thompson@soton.ac.uk)
Stephen V Faraone (sfaraone@childpsychresearch.org)
Hans-Christoph Steinhausen (steinh@kjpds.uzh.ch)

Version: 2 Date: 1 December 2010

Author's response to reviews: see over
To: BioMed Central Editors

November 30, 2010

Ueli C. Müller, PHD
University of Zurich
Department of Child- and Adolescent Psychiatry
8032 Zürich

Dear Editors

We are submitting the revised version of the second of two companion manuscripts entitled

The impact of study design and diagnostic approach in a large multi-centre ADHD study: Part 2: Dimensional measures of psychopathology and intelligence.
(Former title: Behavioural patterns of ADHD in a large multicentre study. Part 2: Dimensional questionnaire data and intelligence measures in probands and unselected siblings of the IMAGE project.)

for your consideration for publishing as a research article in BMC Psychiatry.

The manuscript is carefully revised by the authors. We think having successfully accounted for all the advices of the reviewers and incorporated their suggestions into the manuscript. If the few well-grounded cases we didn’t follow their suggestions, or in cases of misunderstandings, we stated our motivation within the reviewers text. Please see our comments to the authors review attached.

We hope having satisfied the reviewers and the editors demands and would be happy if the manuscript would be published in your journal

Sincerely yours
Review of the paper titled: Behavioural patterns of ADHD in a large multicentre study. Part 2: Dimensional questionnaire data and intelligence measures in probands and unselected siblings of the IMAGE project.

In this manuscript the authors present the behavioural phenotype of children who participated in the International Multicentre ADHD Genetics (IMAGE) project. The authors report data for 1068 probands and 1446 siblings. This manuscript is the second part of the two part submission.

The authors reported significant differences between probands and siblings, and status (proband/sibling) by gender interaction effect on Conners teacher and parent rating. Centre effects were found for teacher ratings and some of the parent ratings. The teacher and parent ratings on Strength and Difficulties Questionnaire differed between probands and siblings on four problems scales, and boys and girls but only for two of the problem scales. Centre effects were also found. Centres also differed on Social Communication questionnaire. On the intelligence measure, IQ was found to be negatively correlated with age, and
significant status, gender and centre effects were found.

This study has several strengths including the sample size and the thorough report of the results. It is well written and gives a good account of the behavioural profile of the sample. Again, inclusion of the teacher as an informant provides an important information about the child’s behaviour at school as well as information about the child as compared to his/her peers.

Suggestions
- In the Methods section the authors describe the sample that includes the 1446 unselected siblings. However, in the Discussion section it is stated that the findings from the questionnaires and intelligence data is described for the probands and unaffected siblings. It would be beneficial to clarify whether siblings with ADHD diagnosis were included in the analysis.

“Unaffected siblings” corrected to “unselected siblings”

- In part one the authors describe the sample and examine siblings who also reach the diagnosis of ADHD. If these children were included in the current analysis, it would be interesting to include the diagnostic status when comparing probands and siblings particularly when ADHD related symptoms and IQ is compared. Identifying the differences and similarities between the affected and non-affected siblings will be very informative for the reader.

Dividing the siblings into an affected and an unaffected group would not be trivial for two reasons.

1) Only a subgroup containing 339 of all 1446 siblings who were clinically rated as “at least suspected to have adhd” based on criteria defined in the study protocol underwent the full diagnostic procedure including the interview. The diagnostic interview for siblings was mainly used to exclude “probands” from the sibling group in genetic analyses. Diagnoses of the remaining 1107 siblings were based on questionnaire data only.
The definition of “affected” is to some extent arbitrary. Probands fulfill the criteria for ADHD-CT (combined type). Should the group of “affected siblings” include all 3 types of ADHD? Or should children with ADHD-CT, i.e. “probands” be excluded? What about subclinical cases?

Because of these reasons we decided to describe differences between (fully) diagnosed subgroups in part 1, and to focus on the differences between ADHD-CT probands and their unselected siblings in part 2. Beyond the description of the sample, the aim of part 2 was to give a more or less naturalistic picture of the unselected sibling population.

- If data is available it would be interesting to see what demographic factors, such as SES or age when children first attend school, can explain the centre effect observed in both papers for most examined measures.

Due to varying definitions in the participating countries there was no common measure for SES. Some proxy measures will be analysed in future studies.

Overall the second part of this submission is also very well written which presents important information about the IMAGE database. I recommend it for publication.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

None
Authors comments in red (30 November 2010)

Reviewer’s report (C. Dolan)

Title: Behavioural patterns of ADHD in a large multicentre study. Part 2: Dimensional questionnaire data and intelligence measures in probands and unselected siblings of the IMAGE project.

The second part concerns questionnaire data (Conners, SDQ, SCQ, WAIS, WISC). The aim of the paper is again to assess the effects of various sources of variation (centre, sex, informant). Again the statistical analysis are all well executed, without making distributional assumptions which are hard to justify. I have some additional comments, some of which also pertain to Part 1.

page 8. I am not familiar with the term "prorated" (near bottom).

This term has a similar meaning like inferred or estimated. It is a quite common term in methodology.

page 15. The country abbreviation are not explained in this paper (other abbreviations are repeated).

Inserted in the methods section.

page 16 and earlier (e.g., page 11). The correlations between age and SDQ scales were weak but significant...... No mentioned is made of the alpha. Is this 0.05? If - .046 is significant at the chosen alpha, the chosen alpha may well be to high given the sample size (power). Is multiple testing an issue here? Given the effect sizes I
suppose that the analysis with or without the age correction produced almost identical results.

A reference to table 3 showing the p-values was added.

Re: multiple testing: again, the main aim is the description of the data and not testing of hypotheses (the same remark was added in the methods section like in paper 1). Here, correlations are calculated in order to decide, whether age should be considered as a covariate or not.

Page 17, and other pages (Part 1 and Part 2). The expression of actual effect sizes in terms of the statistic Q (page 17 line 7 ff), or in terms of actual raw means or difference in raw means is inconvenient to the reader who is not familiar with the particular statistics (Q) or does not have standard deviation at hand. Is it not possible to express effect sizes in terms of some familiar effects sizes?

Effect sizes are indeed reported in a common way (in units of standard deviations, see “statistical procedures” and table 1). In the paper, the term “effect” does not refer to effect size, but is used in a more simple way in terms of the impact of a variable of factor on a dependent variable. Q is not an effect size measure, but a test statistic in multi-way procedures (similar to “t” or “chi²”), see also methods section.

Page 19 bottom. At the risk of repeating myself (this pertains to Part 1 and 2): the Gender x Informant interaction (variable PB) is judged to be significant (p=.043) as shown in Table 4. But in Table 4 there are 24 p. values, all tested individually. It is up to the researcher to determine his/her alpha, but it is useful to state clearly one’s policy w.r.t. alpha. One could argue that 24 test should be tested one-at-a-time using alpha=.05/24 = .002 (say; assuming 0.05 is OK given the sample size and possible
power). In that case only 5 effects are significant. Of course there are other methods of controlling over alpha. This remark obviously also concerns the other tables. Table 2 contains 52 p-values! etc.

See my comments on multiple testing in a descriptive context.

Correction for multiple testing is important, when there is a focus on an generalizing perspective (e.g. “Is there an influence of age on IQ?”) and when positive partial result(s) (e.g. significant correlation between age and the “picture completion” task) are taken as a statistica argument for answering the general question. However, this is not the case in the present paper(s).

Table 4 and 5. Whence the variation in conveying p value (actual values in 4; actual values and *, **, and *** in table 5)?

The differences stem from different outputs of the statistical procedures used in this paper.

Page 23. near bottom. "In contrast, for the Conners’ scales and the IQ scores, age and gender effects, first of all, indicate differences in the expected deviations from normality”. What does this mean, what does "differences in the expected deviations" mean? what expected deviations from normality? Please rephrase for clarity.

Done

Page 24 bottom is nonlinearity is the source of the lower correlations (hyperactivity, in attention), then this could be verified by regressing the variables on age and age-squared. (age should be centered first).

Yes, this could be done. However, we are not dealing with one nonlinear effect between two variables (e.g. between age and hyperactivity), but rather with a
combination of two different symptom clusters with individual developmental curves in one single variable. Finding an optimal function between age and this combination core would solve the problem only partially. Furthermore, our strategy consisted of the use of the same robust statistics across variables to

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I have no competing interests
Reviewer's report:

Overall assessment: The authors are courageous to undertake this investigation of factors that might potentially contribute to the marked heterogeneity of ADHD, particularly because it highlights the grey-zones of diagnosis of mental health conditions. That is, although the rationale is based on potential problems of diagnostic procedures and tools for genetic analyses, the clinical implications are rather concerning. This second manuscript focuses on a different subset of data from the IMAGE project from those used for the first manuscript (i.e., parent and teacher questionnaires analyzed at subscale level, IQ test, as opposed to parent interview and parent and teacher questionnaires analyzed at the item level). However, the analyses herein also examine the same factors as used in the first manuscript (age, gender, informant, centre, tools). The major results indicate differential effects of the proposed factors (age, gender, child-status (proband, sibling), informant, instrument, and diagnostic centre) on the questionnaire-based and IQ data (small gender and status effects on IQ scores, but large effects on questionnaire scores). The results are interpreted primarily from the perspective of the resultant risk for quantitative genetic analyses based on these standardized diagnostic measures. Major strengths of this study on which the analyses are based, include the large sample of probands and siblings amassed across 8 countries, the use of standardized
inclusion/exclusion criteria and the same assessment tools. Of course, the present data set will also reflect any shortcomings in the design or implementation of the IMAGE study itself (e.g., marked differences in sample sizes contributed by the various diagnostic centres/countries; use of assessment tools that may not have been fully adjusted and validated for cross-cultural differences).

Shortcomings of this second manuscript include lack of clarity of the primary aims of the study, the confounding of ‘diagnostic centres’ and countries, the lack of attention to the clinical implications, and the complex statistical analyses (although seemingly robust) will render the results inaccessible to much of the readership of the Journal. Also, in its present state, the manuscript cannot stand alone (ie, without the first manuscript) because key information about the diagnostic procedure and assessment tools are not included herein. Thus one key issue is to what extent the two manuscripts need to be independent of each other and whether two manuscripts are warranted on this topic and justified by the breadth of diagnostic behavioral data generated by the IMAGE study and the importance of the findings.

Major Compulsory Revisions

The major aims of the study need to be clarified and written consistently in the Abstract, Introduction, and Discussion. For instance, the overall aim was implied in the Discussion, by the sentence..”AN important question was whether the gain in power due to the large sample size was at least partly lost by [inadvertently] increasing heterogeneity of the sample.” The specific aim was then stated as being to determine possible sources of heterogeneity. Revised.

2. The rationale for the study presented in the Introduction must be clarified. Reflecting on both manuscripts, this issue has wide-spread implications that go beyond the concerns for molecular genetic analyses. That is, there is increasing pressure from funding agencies to engage in multisite
large-N studies, with the implicit assumption that larger is better. To me, the current set of analyses put this assumption under the microscope – at least for child mental health field. Revised.

3. At present, the Introduction is not well-organized and the underlying logic for the current set of analyses is not clearly stated. By contrast to manuscript #1, the introduction is too short and the reader arrives at the methods section without a solid understanding of the issues to be investigated or of the rationale for the approach. For example, it is unclear why IQ scores are included along with behavior questionnaires, and why the SCQ is included. From a design perspective, the inclusion of IQ measures are advantageous for several reasons: By contrast to the questionnaires, IQ tests are administered to the child under controlled and standardized conditions by a trained psychometrist; performance is scored using strict criteria that minimize subjectivity; the scores index the child’s actual performance that day as opposed to perceived behavior over the past months as rated by significant others; and are normed for age and thus would not be expected to be influenced by age or informant (parent, teacher) etc. Revised.

4. As in the first manuscript, the issue of ‘centre’ versus ‘country/culture’ requires clarification in both the text and analyses. If I understand correctly, two countries contribute samples from 2 diagnostic centres (Netherlands, Germany), whereas other countries contribute participants from just one centre. Moreover, Netherlands contribute 765 participants which is the largest sample from a country (30% of the total sample). From a design perspective, the predominant unit would appear to be best described as country rather than centre. This issue poses problems for interpretation of findings., because it is not at all clear whether ‘centre-based’ differences reflect possible cultural differences or centre-based issues.

The reason for choosing ‘centre’ as recruitment unit (considerable differences in several demographic and psychopathologic measures between different centres of the same countries) is more clearly stated; a reference to corresponding results in paper 1 is added.
5. The Discussion needs to be drastically shortened and better focused. It is currently 10 pages!

Revised.

6. Clinical implications must be discussed. For example, how should clinicians and researchers view the Conners versus the SDQ in terms of clinical and research use? Are the contrast effects important in terms of clinical diagnosis when there are twins or siblings involved?

Revised.

Minor Essential Revisions

1. The whole manuscript will benefit from rigorous editing to correct some understandable problems with English written expression and incorrect statements (e.g., Introduction “Furthermore, intellectual abilities often are impaired in ADHD.” – by definition IQ scores must be in the ‘normal’ range – not impaired as in Intellectual Disability. Research often indicates that IQ scores are a few points lower than controls but they are still within the normal range). Revised by native speaking Co-Authors.

2. The methods section needs to include all of the essential information to understand the study design. The study design, diagnostic protocol and diagnostic procedure should be summarized briefly, with a cross-reference to more detailed descriptions in another manuscript (if appropriate). Revised.

Under Measures, specify the number of probands, siblings to have the WAIS (versus the WISC). Also a comment is warranted on the validation of translated versions of the questionnaires and WISC/WAIS in the participating countries. Revised.
3. The diagnostic protocol requires some clarification. Under “Diagnostic protocol for probands” the authors state that the proband must meet DSM-IV criteria based on CTRS-R:L and the PACS. The CTRS threshold (T-score). There were no T-scores used, but symptom checklists according to the DSM-IV criteria extracted from the Conners’ questionnaires. and specific subscale(s) need to be specified. Also, clarify why pervasiveness was based on the threshold of “2” symptoms on CTRS and without any mention of impairment. Could the 2 symptoms on the CTRS be unique – that is, not endorsed on the PACS interview? See comments and revisions in paper 1.

Discretionary revisions

1. Results: report the male: female ratio is a more standard format (eg. 7.2: 1, instead of 938:130). Clarify to which tables figures you are referring (sometimes difficult to tell whether the table is in additional material or in the manuscript (eg Table 1a, doesn’t seem to exist but Table A1 does and is part of additional material) Corrected.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

I declare that I have no competing interests.