Reviewer’s report

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.

Version: 1 Date: 12 August 2010

Reviewer: conor dolan

Reviewer’s report:

Review by CVDolan of Part 1 and 2 of Behavioural patterns of ADHD in a large multicentre study.
Part 1: Demographics, diagnoses, adn symptoms in probands and selected siblings of the IMAGE project. Müller et al.
The aim of the paper is to describe the IMAGE sample, and to describe and analyze symptom patterns in the probands, concentrating on the emphasis is on ADHD-CT (combined type). As such there is little wrong with this paper: this paper is provides useful and detailed documentation of the IMAGE project, esp. as the data will be made available. Anyone analyzing the data in the future, will want to have this info. at his/her disposal. The statistical analysis are all well executed, without making distributional assumptions which are hard to justify. I have only minor comments, which, in so far as the authors judge them worthy of action, should pose no problem.
The reason for focussing on CT in the probands is not clear. (In the siblings all three types + no diagnosis are considered). I understand that CT may be the disorder of interest in the subsequent genetic analyses, but the paper merely describes the samples + relevant sample effects.
Page 4-6 why heritabilities should constitute necessary or sufficient evidence for the continuous nature of ADHD eludes me.
The Diagnostic protocol on page 8 could be presented a bit clearer. I understand that the diagnosis of ADHD-CT is based on the CTRS-R:L and the PACS. The criteria in terms of items endorsement are given. The role of additional criteria is unclear, specifically clinical impairment and pervasiveness. The criteria for the former is not stated clearly. Clinical impairment was inferred by the presence of ADHD items in the PACS interview.... Do you mean endorsed items?
Page 14 1068 probands and 1446 siblings were included. A chi2=57.1 is
reported, but testing what?

Page 14 (near bottom). Perhaps I missed it but what does “family status” refer to? was that mentioned?

It would be useful to state the policy with respect to multiple testing more clearly. E.g., in the analysis on page 17 effects of gender and centre are studied (on age adjusted onset of inatt. symptoms). The interaction counts as significant, the p value is .028. (See also page 19 near bottom: p=.045; but on page 23, p=.023 is not significant). So the main effects and interactions all tested using alpha=0.05? Also in the post hoc analysis 12 of the 55 tests are significant. Given alpha=0.05 with corrected for multiple testing?

p. 19 near bottom re: p=.045. The effect size is apparently being based on the p value. Is that justified? A small effect can have a very small p value, a large effect can have a p value that is just below the alpha.

p. 34. The issue of outliers is mentioned here and in Part 2 (only in the conclusion). But outliers are not actually assessed are they?

Part 2: Dimensional questionnaire data and intelligence measures in probands and unselected siblings of the IMAGE project. Müller et al.

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).

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

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 (power0. 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.

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?
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 $\alpha = 0.05/24 = 0.002$ (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.
Table 4 and 5. Whence the variation in conveying p value (actual values in 4; actual values and *, **, and *** in table 5)?
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.
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).

**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.