Reviewer's report

Title: Data driven subtypes of major depressive disorder

Version: 1 Date: 12 August 2012

Reviewer: Naomi Wray

Reviewer's report:

This is a useful and thoughtfully conducted study.

Major

1) The authors seem unaware of the latent class analysis conducted on the PGC-MDD study (>9000 cases) (see supplement for details). Most notable in that study was the huge heterogeneity of endorsement of MDD items. This is a key point for paragraph 2 of the discussion.

2) How many studies were excluded on the > 75% need MDD criterion. Why was this criterion needed? I would have thought that non-MDD would fall out as a class because of low endorsement rates. Is this the reason why did the Sullivan et al (AJP 1998) paper did not get included in the analysis? I believe this decision should be reversed unless the authors can provide strong justification.

3) The discussion should include guidelines for best practice for future analysis of MDD criteria data. Eg latent class analysis seems less useful than latent factor analysis if it simply groups on severity.

4) The NIMH is developing Research Domain Criteria (RDoC) is there anything to say from this study to inform this effort?

Minor

1. Table 1 define sx
2. I haven’t gone back to the original studies, but will it be totally transparent which class is a and which is b from each study. I suggest this is included in Supplementary.
3. It is not clear to me that a and b are compatible across studies
4. Put references in the tables

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare I have no competing interests