Author's response to reviews

Title: Data driven subtypes of major depressive disorder

Authors:

Hanna M van Loo (h.van.loo@umcg.nl)
Peter de Jonge (peter.de.jonge@umcg.nl)
Jan-Willem Romeijn (j.w.romeijn@rug.nl)
Ronald C Kessler (kessler@hcp.med.harvard.edu)
Robert A Schoevers (r.a.schoevers@umcg.nl)

Version: 2 Date: 30 August 2012

Author's response to reviews: see over
Dear Dr. D’Souza,

Thank you for sending us the thoughtful editorial and reviewers’ comments on our manuscript “Data driven subtypes of major depressive disorder” and for your willingness to consider a revised version for publication in BMC Medicine. We are glad to see our paper has been read carefully and we appreciate the helpful suggestions of you and the reviewers.

In this letter, we have listed the editorial and reviewers’ concerns. Our responses to those concerns have been italicized and the changes in the text of the review are in bold type. Furthermore, we revised the initial manuscript (changes again in bold type) and added requested information in the Additional files.

We hope we have answered the reviewers’ comments satisfactorily in our revised manuscript. If there are any questions, we are willing to answer them.

Best regards,
On behalf of the authors,

Hanna van Loo

Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE)
Department of Psychiatry
UMCG University of Groningen
Building 24 CC72
P.O.Box 30 001 9700 RB Groningen
The Netherlands
Tel: +003150 3619005/3611242
Fax: +003150 3619722
Responses to the editorial suggestions:
1) Please remove the table of contents as this does not conform to the journal style (http://www.biomedcentral.com/info/ifora/medicine_journals). It is important that your files are correctly formatted.

*We removed the table of contents.*

2) Please ensure your manuscript adheres to MOOSE guidelines (http://www.mat.or.th/journal/download/MOOSE%20Guidelines%20for%20Meta-Analyses%20and%20Systematic%20Reviews%20of%20Observational%20Studies.pdf). For instance, please modify your title to reflect that this is a systematic review.

*We checked the MOOSE guidelines and adapted the review according to those guidelines. First, we adapted the title into: “Data driven subtypes of major depressive disorder: a systematic review”. Second, we provided insight in the qualifications of the Searchers in the Authors’ contributions and Authors’ information. Third, we added a paragraph in the Discussion with Guidelines for future research (p.21), which was also recommended by reviewer 3 and 4.*

3) Please add Authors Contributions and Acknowledgements between your conclusions and references:

*We added Authors’ Contributions between the conclusions and the references. We do not have any further Acknowledgements.*

Responses to the report of Reviewer 1 Professor Michael Berk:
1) This paper was well written, clear and refreshingly free of typographical or grammatical issues. It explores an important issue, particularly in the light of the impending revisions to current diagnostic systems. As the authors note, the existence or otherwise of valid sub-types of depressive disorders is of substantial clinical significance. As such this paper is to be welcomed.

The paper is methodologically rigorous, conservative and thorough. I have very few substantive concerns regarding the methodology employed.

*We thank the reviewer for his positive comments on the manuscript.*

2) One personal grouch: pretty much every article on any topic opens as this one does, with a disconnected and anodyne blurb on prevalence and disability. This is not germane to his paper and the first four sentences, up to reference 8, can be deleted.

*We deleted the first four sentences up to reference 8 and condensed into one introductory sentence: “Major depressive disorder (MDD) is one of the grand challenges in global mental health [1, 2]. In research, a continuous challenge is the diversity in symptoms and pathophysiology of patients classified as having the disorder.”*

3) One of the methodological factors that hinder the discovery of latent sub-types is the primary reliance on rating scales, which were designed to be sensitive to change, as opposed to capturing the detailed phenomenological picture of a disorder. There are, as the authors note, a myriad of phenomena and
co-morbidities that form part of the symptomatic orbit of depression. Instruments designed to capture such phenomenology need to be applied and used prospectively. I agree with the authors’ comments about the importance of combining latent factor and latent class models. One of the other limitations of this data set is that medication is often not controlled for. Being on, or not on, a benzodiazepine or an atypical, for example, will change phenomenology, such is their purpose. Ideally, such studies need to be done on drug-free individuals.

In response to the comments of the reviewer we adapted paragraph 3 in the Discussion (p.19) to address his concerns about the influence of medication and rating scales:

“One possible strategy to improve study methods is data enrichment, since it is clear that the quality of data crucially determines the quality of study outcomes. First, some of the studies reviewed above indicate the possible benefit of dynamic measurements as they showed that differentiated symptom profiles might be clearer with more versus less severe cases or early in the treatment process than later in the treatment process. Therefore it would seem prudent to study changes in symptom structure and severity over the course of treatment, preferably also distinguishing the influence of medication. A second set of choices would involve the scales used to assess these symptoms. Some diagnostic instruments used yes-no dichotomous measures for each symptom, while others have gradual assessments. Gradual assessments, of course, would be expected to provide for textured differentiation and in this way possibly lead to greater precision in detecting meaningful subtypes. Another concern is that the frequently used rating scales are primarily designed to be sensitive to change as opposed to capturing the detailed phenomenological picture of MDD. To address the potential heterogeneity in depressive patients at least all DSM criteria should be measured standardly, including all disaggregated somatic symptoms (s3-s6). A third possibility for data enrichment is the inclusion of other variables in addition to depressive symptoms in analyses. It seems clear that the set of symptoms to be included in latent variable analyses should be broad so as to allow for the possibility of detecting subtypes associated with symptoms beyond those in the current DSM and ICD systems. For example, there is some suggestion in the literature that there might be value in differentiating irritable versus non-irritable MDD [59] and to assess the influence of anxiety [60, 61].”

Responses to the report of Reviewer 2 Professor Amanda Baker:
This paper on data driven subtypes of major depressive disorder involved a systematic review of published studies. The methods were well described and the results are of major interest to clinicians and researchers.

We thank the reviewer for her positive comments on the manuscript.

Responses to the report of Reviewer 3 Professor Tim Croudace:
I agree with the four sentences that provide the Background for this paper. I think it is definitely true that bottom-up approaches from signs and symptoms are warranted in the main-stay of psychiatric research. This angle characterizes much of the motivation behind my own work and my motivation in this area.
It seems sensible to attempt a systematic review, to ASSEMBLE papers that have applied different techniques to relevant data. This paper succeeds in this respect for one domain of morbidity (depression criteria of DSM-IV). Although I think it is useful to document what studies were found and the types of
(latent structure) analysis that were performed, it remains a difficult issue for me to accept that comparing results across studies in this manner provides enough useful information. I think this is really an area where, within each study, with high quality data, more can be learned when a particular study compares and contrasts methods on the same dataset. I only wish more literature accepted that challenge, rather than pushing a set of results from one modelling perspective. I think this paper is of interest therefore, but ultimately slightly flawed in the ability to succeed in its aims. I enjoyed reading it, and suspect others will too, but I think that the results reported are just an arbitrary set of those published, from amongst those that could have been performed. We need more papers applying more types of analyses to the same data, in my view. If this paper were to achieve that, then it would be a valued contribution, changing the nature of some of the to-be-published work in this field in the future. The most challenging area is how to appraise the more complex "hybrid" type models, and also to work with different measurement levels for the original symptom ratings/responses. This represents a thoughtful account of important avenues for research though, and offers some accurate arguments and perspectives for a general medical readership (of this journal). I think that the conclusions are accurate and fair.

Based on the results of our review, we fully agree to the suggestion of this reviewer that a reconsideration of study methods and models is desirable. We started our review with the question whether previous studies have showed a symptomatic subtype of MDD and we were not able to answer this question conclusively. The results of the collected studies were too diverse to provide adequate evidence for the existence of qualitatively different subtypes or symptom dimensions of MDD, which we describe in our discussion and conclusion. Together with the reviewer we hope that our review stimulates researchers in the direction of applying and testing more and different types of statistical analyses to the same data instead of adhering to one modeling perspective. We address this in our Discussion in paragraph 4, 5 and 6, which is a new paragraph concerning future analyses.

Major Compulsory Revisions
1 Could some more suggestions be made as to how to address patient heterogeneity in somatic symptoms.

One of the problems in the assessment of the heterogeneity of somatic symptoms is that only a few questionnaires measure all somatic symptoms (see “Measurement of depressive symptoms”, Results p.9). We think at least a minimal amount of somatic symptoms has to be measured to enable the discovery of somatic heterogeneity, if there is such. In response to this question, we have changed the Discussion, paragraph 4 (for the full paragraph, see the answer to Reviewer 1, point 3): “To address the potential heterogeneity in depressive patients at least all DSM criteria should be measured standardly, including all disaggregated somatic symptoms (s3-s6).”

2 Could the authors forecast whether it is ever likely that a mega-analysis could be performed using IRT linking of some of the questionnaires, for example.

This is an interesting point that has also been raised by other reviewers. In answer to the general comments, point 2 and 6 of Reviewer 3, and point 3 and 4 of Reviewer 4, we included a new paragraph in the Discussion with recommendations for Future analyses (Paragraph 6, penultimate paragraph, starting after: “However, if subtypes of MDD-patients exist, taking different statistical methods to reveal their structure could be worthwhile”, p.21):
“Thus, future analyses ideally explore several advanced statistical techniques on enriched data sets. An investigation of the possibilities and limitations of different modeling techniques seems more reasonable than adhering exclusively to the latent factor and latent class models used up to now. Mega-analyses of the MDD-symptoms of different samples could be worthwhile as combined data may yield profits for the robustness and generalizability of the results [55, 68]. Nevertheless, when performing mega-analyses it is even more important to have rich data sets and to apply sophisticated modeling techniques to the data to accommodate inter-study heterogeneity [55]. Experiences with those new symptom based classification attempts might inform other data-driven classification attempts which go beyond the DSM, such as the Research Domain Criteria [69, 70].”

3 Overall, I think that the paper needs to reflect both factor analysis terminology and that of item response theory, since they are essentially the same model. Is enough mention made of multidimensional item response theory, or mixture models (IRT-based)?

We now refer as well to models based on item response theory, in paragraph 4 of the Discussion (see p.20, reference 64, Tay 2011). Furthermore we included a reference to Cole 2011 (suggested by this reviewer, see point 6), who also uses item response theory to assess depression in children.

“A second strategy to improve study methods involves the statistical approaches used to uncover depressive dimensions and subtypes. In addition to the latent factor and latent class approaches, it is possible to consider complex cluster analysis models, especially those that use a canonical formulation to predict diverse outcomes [62, 63] and mixture models that combine features of latent class and item response theory models [64] or latent class and latent factor models [20]. Some recent studies have used factor mixture analyses to identify subtypes in other psychiatric disorders, such as attention-deficit/hyperactivity disorder [65], posttraumatic stress disorder [13] and schizophrenia [66]. So far, these approaches have not yet been used to search for subtypes of MDD but are attractive alternatives in light of their success in detecting useful subtypes of other disorders. Obviously, in accordance with the described latent variable models, many theoretical aspects should be considered before applying those new techniques [67]. However, if subtypes of MDD-patients exist, taking different statistical methods to reveal their structure could be worthwhile.”

4 It is very important to specify exactly which studies used binary and ordinal data as input to their analyses (vs any continuous).

We added in the Results section “Measurement of depressive symptoms” the following sentence (p.9): “Answer categories were binary in 3 studies (yes-no dichotomized), ordinal in 15 studies and mixed in 2 studies.” Furthermore, we added a row in the Study Characteristics table where we indicated the structure of the items for each study (see additional files).

Minor Essential Revisions
5 The introduction does not really give any representation to the literature(s) on discrete latent trait models, or latent class models with random effects - both of which might be relevant.

In the Introduction we describe the two latent variable models which have been used in the studies we identified, i.e. latent factor and latent class models. Although it was not the intention to provide a full
overview of all current statistical models available to discern subtypes, one of the sentences in the introduction might have given this impression, viz. “Two types of latent variable models can be distinguished: latent factor models and latent class models [20]”. To take away this suggestion we adapted this sentence: “Two dominant types of latent variable models are latent factor models and latent class models [20]”.

Discretionary Revisions

6 The authors might like to consider related perspectives published in articles by David Cole et al. (for adolescent depression).

We have included a reference to Cole 2011 which describes a mega-analysis of MDD-symptoms in children in the new paragraph on future analyses (p.21, see answer on Reviewer 3, point 2).

7 It is likely that related studies outside of this systematic review remit are actually relevant to the direction in which the authors would like the field to travel. If some indication of those could be introduced, then I think it would make the discussion more rounded (and inclusive), in some way.

In response to the reviewer, we added a reference to a recent study using a IRT-based mixture model (multilevel mixed-measurement IRT analysis) to investigate the latent heterogeneity in self-reported emotions of > 100,000 individuals from 116 countries in paragraph 4 of the Discussion on new statistical approaches (p.20, reference 64, Tay 2011). Furthermore, we refer to some promising studies outside this systematic review. We mention for instance studies which try to subtype other psychiatric disorders (schizophrenia, ADHD, PTSD) using factor mixture models.

Responses to the report of Reviewer 4 Professor Naomi Wray:

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.

We thank the reviewer for her positive comments and for suggesting the PGC-MDD study of the Psychiatric GWAS Consortium. This study describes a mega-analysis of genome-wide association studies, and also includes a table with MDD-criteria endorsement rates per included study (Table S12. MDD A criteria endorsement rates by study). It shows striking differences in MDD-symptom endorsement between studies. We expect that the studies we identified will also show such differences in symptom endorsement, if only for the fact that the several studies reported varying severity scores of their patient samples. The PGC-MDD study provides empirical support for the suggestion that the severity and quantity of symptoms of the respective patient samples might differ and thus affect the resulting latent classes and factors. Therefore, we added the following sentence (paragraph 2 of the Discussion, p.18):

“How can we explain this great diversity? From the collected results, it can be observed directly that all sorts of factors influence the outcomes of latent variable analyses. For instance, the included number of patients, the severity and quantity of their symptoms obviously affect the resulting latent classes and
Presumably, there is considerable difference in symptom endorsement rates between the studies, as has been observed in a recent mega-analysis of genome-wide association studies of MDD-patients [55].”

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.

Our aim was to search for heterogeneity within the group of patients with MDD, therefore we initially preferred studies with MDD-patients only. However, we found that a substantial part of the studies satisfying all other inclusion criteria included a small percentage of patients with minor depression, adjustment disorder or dysthymia (7 out of 20 included studies, see Study Characteristics for percentages). We considered the presence of those patient categories in small percentages (<25%) acceptable. However, we excluded the studies including more than 25% of those patient categories in order not to add too much heterogeneity to our intended patient group. Thus, to provide a more complete overview of the literature we chose to describe those studies as well. We agree that the percentage is arbitrary.

The study of Sullivan et al. (Am J Psychiatry 1998; 155:1398–1406) was excluded since there was no diagnosis of MDD according to the DSM (“To be included in the latent class analysis (LCA) in our present study, the worst lifetime episodes from the NCS (N=2,836) must have lasted at least 2 weeks, been associated with help-seeking or impairment, and contained one or more contemporaneous depressive symptoms. These depressive syndromes may or may not have met DSM-III-R criteria for major depression” P.1399).

To better explain the use of this criterion we added the following sentence to “Inclusion criteria” (p.7): “We did not use a stricter criterion of 100% MDD-patients, since we did not want to exclude studies with a minor percentage of patients with minor depression, adjustment disorder and dysthymia for reasons of completeness.”

In the Results, Literature search (p.9) we added: “Overall, 96.6% of the study subjects satisfied the criteria for major depressive disorder. A small minority (3.4%) of patients suffered from other diagnoses such as minor depressive disorder or adjustment disorder, which were found in 7 out of 20 analyses (see additional files).”

3) The discussion should include guidelines for best practice for future analysis of MDD criteria data. E.g. 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?

In answer to the general comments, point 2 and 6 of Reviewer 3, and point 3 and 4 of Reviewer 4, we included a new paragraph in the Discussion with recommendations for Future analyses (Discussion Paragraph 6, p.21).
Minor

1. Table 1 define sx.

We substituted “sx” by the word “symptom” in Table 1.

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.

In answer to points 2 and 3:

We included an extra table in the Additional files: “Latent class analyses: translation of original clusters to severity based classes”. In this table we show the clusters which were originally identified by the latent class analyses, the number of patients in each original cluster and how they are recoded in a,b,c,d,e based on severity scores. In answer to point 3 we included the sum score of the original clusters on the complete questionnaires in this table to provide insight in the comparability and differences between the respective classes based on severity.

4. Put references in the tables.

We added the references of the included studies as an additional file.