Author's response to reviews

Title: Major Depressive Disorder as a Nonlinear Dynamic System: Bimodality in the Frequency Distribution of Depressive Symptoms over Time

Authors:

Bettina Hosenfeld (b.hosenfeld@umcg.nl)
Elisabeth H Bos (elske.bos@med.umcg.nl)
Klaas J Wardenaar (k.j.wardenaar@umcg.nl)
Henk Jan Conradi (H.J.Conradi@uva.nl)
Han LJ van der Maas (H.L.J.vanderMaas@uva.nl)
Ingmar Visser (i.visser@uva.nl)
Peter de Jonge (Peter.de.Jonge@umcg.nl)

Version: 3
Date: 10 June 2015

Author's response to reviews: see over
Dear Dr Clark,

Thank you for considering a revised version of our manuscript “Major Depressive Disorder as a Nonlinear Dynamic System: Bimodality in the Frequency Distribution of Depressive Symptoms over Time”.

We thank the reviewers for their thoughtful comments. We revised the manuscript and addressed each comment raised by the reviewers. Below is indicated how the comments were taken into account in the revised version. We hope that you will now find the manuscript suitable for publication in your journal. Please let us know if any other changes are needed. We look forward to hearing from you.

On behalf of all authors, sincerely,

Peter de Jonge
Reviewer: William Lee

Reviewer's report:

This is an excellent paper which deserves wide readership. The authors used repeated measures to examine the course of symptoms of depression. My only comments are small and primarily of the 'suggestion' or 'no for publication' variety:

1) Abstract>Methods: "recorded weekly during years" is not very clear or very good English. "Recorded weekly for two years" might be more suitable.

According to the reviewer's suggestions we made the following change in the text:

‘...In 178 primary care patients with MDD, the presence of the nine DSM-IV symptoms of depression was recorded weekly for two years....'(pp 2).

2) Abstract>Conclusions: "their course data" might be better expressed as "life course data"

We changed this phrase according to the reviewer's suggestion:

‘...The BC seems useful in differentiating between subgroups of MDD patients based on their life course data...’ (pp 2).

3) Last part of Introduction, para 1: cf. [9] could be better expressed. A sentence or two outlining the referenced paper and why it should be compared with the subject under discussion would help to set the scene for the general psychiatric reader.

In the revised manuscript, we provided a short summary of reference 9:

‘...This hypothesis has already been proposed by several authors [5,6,7,8,9], but has not yet been systematically investigated empirically. In one study [9], critical slowing down as an indicator of nearby tipping points predicted mood shifts in depressed patients. This finding provides some indirect evidence for the presence of alternating stable states in depression' (pp. 3)

4) Methods>participants: What is the justification for accepting 67% of records with a cut off of 104 weeks? Presumably more weeks would have meant fewer participants and vice versa. I would like to see some justification for why the cut-offs were used (or a simple assurance that this choice was not made after looking at the data.

Both reviewers mentioned some concern about the data loss in the longitudinal study and asked how we decided which cases to include. Before conducting the main analyses, we inspected the frequency distributions and chose the cut-off of 104 weeks (two years), which resulted in 178 complete cases. Because many participants discontinued providing symptom scores after the first two years of the study, this decision seemed to result in the most optimal trade-off between the number of included participants and the amount of completed measurements (see Figure R1). If we had kept the data of only the participants who continuously provided symptom scores for three years, our sample would have shrunk to 121 complete cases (45.3% of the sample).

We clarified this in the Methods section:
Of the 267 participants, 178 patients had complete symptom records for a 2-year follow-up period (104 weeks; 67% of total sample) were obtained and formed the basis for the analyses reported below. Although some patients were followed for a longer period (up to 3 years), the cut-off of 104 weeks (two years) was chosen after inspection of the frequency distributions of responders on each of the time-points. This showed that this cut-off of 104 weeks, 178 complete cases (66.7%) could be included, which formed a more optimal trade-off between sample-size and follow-up time than when, for instance, using a cut-off of 156 weeks, which would have resulted in inclusion of only 121 participants (45.3%).

Figure R1: Number of participants who provided symptom scores from week 1-week 156 (left) and the number of drop-outs at each of the follow-up points (right)

5) Results: time series plots: Some reference to how these groups were defined is appropriate. Same with histogram section. It seems this was simply from observation.

We agree that this could be clearer and added the following paragraph to the Methods section:

‘We inspected the time series plots broadly, applying the three categories “fluctuating symptoms around a mean”, “continuous decrease and/ or increase of the symptoms”, “clear sudden jumps visible”. Similarly, we categorised the histograms as either “unimodal with a mode at the middle of the scale” or as “unimodal with a mode at one end of the scale” or as “bimodal”. The systematic coding of the plots was conducted by the first author. The time series plots were coded independently from the histograms. No other information was used for coding.’ (pp. 8).

6) Discussion>limitations: "Offset" does not mean the opposite of onset, and is best not used here.

We agree and changed the text as follows:

‘Possibly, the participants of the study reported the onset and remission of the symptoms as more or less simultaneously only because it is easier to recall and to report simple patterns than more differentiated or random ones.’ (pp. 16).
Reviewer: Sean Lynch

Reviewer’s report:

I thought overall that this was a stimulating and well-written paper. It is certainly innovative in approach and could be useful in leading to greater understanding of recovery trajectory and normal time course of depressive episodes.

1. The strengths are that the cohort studied is selected from primary care as far as I can understand and is a reasonable size for an initial “proof-of-concept” study. The sample is thus less likely to be treatment resistant or have complex comorbidity. Data has been collected on mood shifts over a two year period for a reasonable proportion of the sample, sufficient to begin exploratory mathematical modelling.

We agree that the used sample was of limited size and also that the study should be seen as a proof-of-principle study. We added some phrases to the manuscript to make this clearer to the readers (see our response to comment 6 for a detailed account).

2. The reviewer noted that …data loss is substantial and needs far fuller discussion.

Indeed, complete data were only available for a subset of the full sample. We included a more detailed description of the selection procedure to make clear why we made the inclusion choices that we made:

‘Of the 267 participants, 178 patients had complete symptom records for a 2-year follow-up period (104 weeks; 67% of total sample) were obtained and formed the basis for the analyses reported below. Although some patients were followed for a longer period (up to 3 years), the cut-off of 104 weeks (two years) was chosen after inspection of the frequency distributions of responders on each of the time-points. This showed that this cut-off of 104 weeks, 178 complete cases (66.7%) could be included, which formed a more optimal trade-off between sample-size and follow-up time than when, for instance, using a cut-off of 156 weeks, which would have resulted in inclusion of only 121 participants (45.3%).’ (pp. 7)

In addition, we included a discussion on data-loss in the limitations section of the manuscript:

‘…we included only complete cases in our analyses. Possibly, the missing values in the original dataset were systematically missing values. Those patients, for example, who did not recover during the intervention, might have been the ones who provided the incomplete records.’ (pp. 17)

3. There are a few issues I have in terms of methodology and other possible limitations of the study. The degree (severity) of depression and number of previous episodes are not defined. Particularly whether this has resulted in self-harm, contact with secondary mental health care services or hospitalisation.

We agree that our manuscript becomes clearer if more information is provided about the baseline clinical status of the sample. We therefore extended Table 1 in the revised manuscript with information that will give the readers better insight into the kind of patients that were included in the study: i.e. depression severity, the number of previous episodes, prior suicidality, comorbidity, the presence of chronic somatic disease, and a number of other clinical characteristics. The following text was added to the manuscript:
The mean Beck Depression Inventory score was 19.7 (SD=9.2), indicating moderate depression severity [24]. The median number of previous episodes was 2 and the mean age of first depression onset was 31.9 years (SD=13.0). Of the patients, 9.6% reported a previous suicide attempt in their lifetime. Percentages comorbid psychiatric disorders in the month prior to baseline ranged from 7.3% for dysthymia to 12.4% for social phobia. The study sample did not differ notably from the original sample (n=267) with respect to socio-demographic and psychiatric characteristics (Table 1).’ (pp. 10)

4. It is not clear whether patients had completed a treatment protocol / programme during entry to the study or whether this study was designed to be more naturalistic.

To make this clearer we added a more detailed description of the inclusion criteria and the randomization procedure.

‘... A total of 397 patients were referred by 49 GP-practices in the North of the Netherlands. Inclusion criteria were: a history of depression, the absence of life-threatening somatic diseases, and receiving no psychotherapy. Patients were excluded if they were pregnant, had dementia, had bipolar disorder, had a psychotic disorder and/or had a primary diagnosis of alcohol or drug dependence. The referred patients were interviewed using the Composite International Diagnostic Interview (CIDI) to confirm the presence of a major depressive episode and absence of other psychopathology. Out of the 397 patients, 78 refused to participate and 52 met the exclusion criteria, resulting in a sample of 267 patients (67.3%). These patients were randomly allocated to four treatment arms: (1) care as usual (CAU) following national general practice guidelines, (2) CAU plus psycho-education program (PEP), (3) CAU + PEP + cognitive behavioural therapy, and (4) CAU + PEP + psychiatric consultation.’ (pp. 6).

5. This leads on to questions as to whether patients continued to have treatment and their degree of response, conventional categorical methods could be used e.g. 50% reduction in initial score (response), 25-50% reduction in initial score (non-response), full remission and response (do not meet diagnostic criteria and low score), which I am sure could be considered by the authors.

We agree that some information on recovery would be informative for the readers. Therefore we added information about recovery (percentages of no, ≥25% and ≥50% reduction in BDI scores) at 6 months and after 1 year after inclusion in the study. The following text was added to the manuscript:

‘...Of the patients with a 6-month BDI follow-up, 21.1% showed an increase or no change in BDI score, 66.2% showed a reduction of at least 25% and 42.1% showed a reduction of at least 50% compared to the baseline BDI-score. Of the patients with a 1-year BDI follow-up, 18.2% showed an increase or no change, 69.7% showed a reduction of at least 25% and 49.2% showed a reduction of at least 50% compared to baseline’. (pp. 10).

5. Other data which I think should be briefly mentioned are any psychiatric co-morbidity (even if not meeting exclusion criteria) or physical co-morbidity, alcohol, smoking and substance use and basic economic and socio-demographic data.

We agree and added this information to Table 1 (see also our response to point 3).

6. As the authors stated, a major methodological weakness is in recall bias due to retrospective recollection of mood score. If this was the most feasible method, possibly a small subgroup could be studied to compare both prospective and retrospective methods of collecting this data? Other options
are to look at modern technologies to instantly upload mood scores. Little data is given on any changes in external life events or stressors or in treatment plan.

The reviewer is right that the retrospective recollection of mood scores is a limitation of our design and provides a useful suggestion. We added the following text to the Discussion:

‘Furthermore, it might be useful to register the symptoms of depression prospectively, at the moment at which they are experienced and even more frequently than only once a week so that the risk of study artefacts will be minimized. Small hand-held devices have already been used in order to record symptoms or mood scores several times a day [33]; similar applications for cell phones might further facilitate registration. A validation study comparing retrospective with prospective recordings of depression symptoms might reveal systematic recall errors.’ (pp. 17-18).

6. The statistics are described in detail, but threats to their interpretation e.g. modest numbers and repeated (possibly related) measures need fuller discussion. I think that for all the above reasons if the authors were to state that this is a preliminary or “proof-of-concept” study, this might be much clearer for the readers of the journal.

We agree that, given its size and inherent limitations, our study should be seen as a proof-of-concept study. Therefore, we added a phrase to the Background section, stating that our study is a “preliminary proof-of-concept study”.

In the Discussion section we now give the following overview of the potential limitations to our data and analytical approach, again stressing that the study should be seen as a proof of principle study:

‘The current study was intended as a preliminary proof-of-principle study and several limitations should be kept in mind when interpreting the presented results. First, it could be that the bimodality pattern we discovered in part reflects artefacts stemming from the data-collection method. All information about the symptoms was recorded retrospectively over three-month periods. Possibly, the participants of the study reported the onset and remission of the symptoms as more or less simultaneously only because it is easier to recall and to report simple patterns than more differentiated or random ones. Alternatively, the patients might have reported a two-state pattern because it fitted their expectations about depression. Three of the four treatments, however, included the instruction to monitor the onset of depression symptoms. All patients who received enhanced treatment were encouraged to keep diaries about the symptoms they experienced and might have consulted these diaries when responding to the interviewers. Accordingly, one would expect that the subgroup of patients receiving care as usual presented the least differentiated symptom curves. Nevertheless, we found no differences between the symptom patterns of the four treatment groups. Second, we included only complete cases in our analyses. Possibly, the missing values in the original dataset were systematically missing values. Those patients, for example, who did not recover during the intervention, might have been the ones who provided the incomplete records. Third, partly due to the selection procedure, the study sample ended up being relatively small. Fourth, the study was conducted among primary care patients and the results may not be directly generalizable to more severely affected patients with potentially more complex comorbidities and/or treatment resistance. Finally, the current analyses were conducted under the rather strict assumption that the symptom sum score could be seen as a unidimensional representation of underlying depression severity and that between-person differences could be quantified in terms of variations on this dimension.’ (pp. 16-17).
7. Although the authors are enthusiastic about the possible bimodality or multi-modality constructs for depression, how can they be certain that these are distinct states i.e. “depressive” and “non-depressive”?

We agree that we cannot be completely sure that our results point toward the general existence of two discrete states of depression (depressed vs. non-depressed). Rather, our results say something about the form that within-person changes in depressive state can take and the heterogeneity that exists across patients. Based on the results, we concluded that in some, levels of depressive symptomatology may vary continuously, whereas in others they may vary between two extremes. As the BC is a continuous measure, many in-between forms can be observed between the two extremes (low BC vs. high BC). This suggests that the dynamics of symptom-levels over time show considerable heterogeneity across patients, which could indicate that different underlying mechanisms are involved.

We clarified this in the Discussion section:

‘...These findings suggest that the dynamics of depression over time differ across patients, with many showing more or less abrupt transitions from one state to the other and others showing more continuous variation in depression severity over time. On the one hand, these results support the traditional all-or-nothing distinction between depression and health. On the other hand, the results show that bimodality is a matter of degree and that in many patients continuous temporal variations in depression severity also play an important role. These variations are a potentially important source of between-person heterogeneity and could indicate the involvement of different underlying mechanisms.’ (pp 14).

8. Also, it is important to consider typology of depressive disorder i.e. interpretation of data could also depend on whether severe and moderate severity forms of depression are seen as within the same spectrum of disorders or as being different disorders, as is the view of some researchers. It is possible that there may be several states e.g. full episode, partial remission, subsyndromal (but not in normal range) and normal mood reactivity states? I would like to see the authors engage in much more detailed discussion of these issues.

This is indeed an important issue. Our results are all based on sum scores of the DSM-IV depression criterion-symptoms. By taking this approach, we implicitly assumed that depression severity is unidimensional, and that differences between persons could be explained by quantitative differences (changes over time) on this severity dimension (i.e. in line with assuming that minor, moderate and severe depression are different levels on the same underlying severity dimension). We agree with the reviewer that patients could alternatively be differentiated from each other in other ways, for instance, by looking at more qualitatively different symptom-profiles, clinical stages, and/or normal vs. pathological reactivity of mood. However, the unidimensional approach of depression was chosen as it was optimal to explore our basic research question using the BC, which requires the use of a single outcome, of which bimodality can be quantified. We agree that other underlying models could offer an even better description of inter-personal differences in depression course dynamics. We added a phrase about this in the limitations:

‘...the current analyses were conducted under the rather strict assumption that the symptom sum score could be seen as a unidimensional representation of underlying depression severity and that between-person differences could be quantified in terms of variations on this dimension.’ (pp 17).
8. The authors should finally discuss how they would like to see research develop e.g. if larger and multi-centre prospective studies.

In the revised manuscript we offer some suggestions for further research:

‘...Ideally, our results should be replicated in larger samples of patients with MDD, by preference in a multi-centre study. Mood recordings from healthy controls might further contribute to insight in the dynamics of depression. Keeping the participants motivated to record their mood continuously and reliably should be a major aspect of the study design.’ (pp17).

9. Language corrections:

Before resubmission, the manuscript was edited by a qualified copy-editor.