

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Efficacy and acceptability of next step treatment strategies in adults with treatment-resistant major depressive disorder: Protocol for systematic review and network meta-analysis
<b>AUTHORS</b>	Muit, Jan; van Eijndhoven, Philip FP; Cipriani, Andrea; Dalhuisen, Iris; van Bronswijk, Suzanne; Furukawa, Toshi; Ruhe, Henricus

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Kessler, David University of Bristol, population health sciences
<b>REVIEW RETURNED</b>	25-Oct-2021

<b>GENERAL COMMENTS</b>	<p>This is an important topic. TRD is very common. The review is timely for a number of reasons: the limitations of previous reviews; the increasing use of newer modalities of treatment eg neuromodulation and psychedelic-assisted psychotherapy. The protocol is clearly written and comprehensive. The eligibility criteria for inclusion of studies are stringent; I wondered if it was really necessary to insist on studies having at least one 'prospective ascertainment of treatment failure'. This may exclude a number of pragmatic or community studies where a record of having been treated and the participants response to adherence questions have been thought to be adequate.</p> <p>I have suggested 'specialist statistical review' simply to ensure that someone who is more expert than me in the methodology of NMA reviews this paper.</p>
-------------------------	---

<b>REVIEWER</b>	Hengartner, Michael Zurich University of Applied Sciences/ZHAW
<b>REVIEW RETURNED</b>	02-Nov-2021

<b>GENERAL COMMENTS</b>	<p>This is a study protocol addressing an important research question. I have only a few minor suggestions.</p> <p>The definition of TRD is highly problematic (failure to respond to more than 1 adequate antidepressant trial). I'm well aware that this is a common approach to define TRD, but it's still grossly inadequate (Fava et al 2020; Psychother Psychosom;89:265-273). According to this definition, about 50% of patients with depression have TRD. The authors should consider using a more stringent definition of TRD, for instance at least 2 failed antidepressant trials, and better yet, 2 failed antidepressant trials and 1 failed psychotherapy trial. This should be elaborated on in the manuscript. At the very least the authors should conduct sensitivity analyses according to these different definitions (by excluding the</p>
-------------------------	--

	<p>less stringent definitions), in addition to the planned meta-regression analyses.</p> <p>The authors should also assess quality of life, if possible.</p> <p>The authors should specifically focus on long-term outcomes (at least 6 months), for acute short-term outcomes (typically 6-8 weeks) are of very limited relevance to patients (and also of limited utility to clinicians).</p> <p>The definition of no treatment (&lt;50% of patients receiving antidepressants) is problematic. According to this definition, patients that received a non-pharmacological treatment (psychological and/or psychosomatic, e.g. exercise) are considered untreated.</p> <p>How do the authors intend to account for varying study duration? Will they report short- and long-term results separately? This could be a serious issue, because I assume that trials testing a psychological intervention last longer, on average, than trials testing a pharmacological or neuromodulation intervention.</p>
--	---

<b>REVIEWER</b>	Xu, Jiaqiong Houston Methodist Research Institute
<b>REVIEW RETURNED</b>	25-Jan-2022

<b>GENERAL COMMENTS</b>	This is a well-written manuscript. The statistical analysis plans are presented clearly and well. I don't have any comments.
-------------------------	--

## VERSION 1 – AUTHOR RESPONSE

### Comments of dr. David Kessler

1. *This is an important topic. TRD is very common. The review is timely for a number of reasons: the limitations of previous reviews; the increasing use of newer modalities of treatment eg neuromodulation and psychedelic-assisted psychotherapy. The protocol is clearly written and comprehensive.*

We thank the reviewer for these kind words, and the acknowledgement of the importance of the research topic.

2. *The eligibility criteria for inclusion of studies are stringent; I wondered if it was really necessary to insist on studies having at least one 'prospective ascertainment of treatment failure'. This may exclude a number of pragmatic or community studies where a record of having been treated and the participants response to adherence questions have been thought to be adequate.*

We apologize for the confusion regarding the topic of eligibility criteria. Indeed, we will not exclude studies without a 'prospective ascertainment of treatment failure'. We do, however, aim to perform a sensitivity analysis with only studies with a prospective ascertainment of at least one treatment trial failure, as described on page 18 of the manuscript:

“Sensitivity of our conclusions for the two primary outcomes will be evaluated by analyzing:

(..)

(4) only studies with a prospective ascertainment of at least one treatment trial failure ”

We also added the following statement to page 9 of the manuscript to clarify this:

“We will include studies with both prospectively and historically assessed treatment failure.”

### **Comments of dr. Michael Hengartner**

1. *This is a study protocol addressing an important research question. I have only a few minor suggestions.*

We thank the reviewer for his kind words, and the acknowledgement of the importance of the research topic.

2. *The definition of TRD is highly problematic (failure to respond to more than 1 adequate antidepressant trial). I'm well aware that this is a common approach to define TRD, but it's still grossly inadequate (Fava et al 2020; Psychother Psychosom;89:265-273). According to this definition, about 50% of patients with depression have TRD. The authors should consider using a more stringent definition of TRD, for instance at least 2 failed antidepressant trials, and better yet, 2 failed antidepressant trials and 1 failed psychotherapy trial. This should be elaborated on in the manuscript. At the very least the authors should conduct sensitivity analyses according to these different definitions (by excluding the less stringent definitions), in addition to the planned meta-regression analyses.*

We thank the reviewer for raising this important issue. We agree TRD definitions are very heterogeneous, as previously described by Berlim and Turecki (2007). Because of this variability, we aim not to be too stringent in the inclusion criteria regarding TRD, because otherwise we would potentially lose studies that might be of interest. See also point #2 raised by dr. Kessler. However, we do aim to investigate this important issue by exploring and operationalising the effect of various definitions of TRD, as described on page 15 of the manuscript:

“In order to answer our third clinical question, we will perform meta-regression that evaluates the impact of different levels of TRD on the primary outcomes. TRD is defined as (i) the number of failed (antidepressant) treatment-trials (including augmentation and psychotherapy) that were required as inclusion criterion for the study or (ii) dichotomized by slightly adapting Conway, 2017: TRD level I (Failure of 1 or 2 adequate dose-duration antidepressants or psychotherapy from different classes (either in combination or succession)) or level II (Failure of ≥3 adequate antidepressant or psychotherapy trials from different classes (either in combination or succession)). If sufficient data are available, we aim to use the first, more detailed, grouping of TRD. If this proves unfeasible, we will employ the second definition.”

In order to further address this issue, we amended the protocol by replacing one of the sensitivity analyses (p18) with:

“only studies that required at least two treatment trial failures in their definition of TRD”.

In order to limit the number of total analyses, we removed

“only studies that used non-response as inclusion criterion (i.e., we exclude studies that used non-remission as an inclusion criterion)”

as a sensitivity analysis. We removed this particular analysis because it was already partially overlapping with the meta-regression analysis for “depression severity at baseline”. The following sensitivity analyses are now described on page 18 of the manuscript:

“Sensitivity of our conclusions for the two primary outcomes will be evaluated by analyzing:

- (1) only studies with reported SD rather than imputed
- (2) only studies that required at least two treatment trial failures in their definition of TRD
- (3) only studies with a low risk of bias
- (4) only studies with a prospective ascertainment of at least one treatment trial failure ”

3. *The authors should also assess quality of life, if possible.*

We thank the reviewer for this good suggestion, and we agree outcomes of trials in MDD should acknowledge long-term quality of life outcomes too instead of short term efficacy only. We however have chosen to abstain on an analysis of quality of life (QoL), because we anticipate to find only a small number of studies that will be consistently reporting this outcome. Previous authors have not been able to meta-analyse data on quality of life in depression due to significant heterogeneity in definitions of QoL (or lack thereof), instruments used to assess QoL and length of follow-up (Sivertsen et al. 2015). This problem in MDD research should be addressed in future trials. We have added this as a limitation of the study, on page 3 under ‘*Strengths and limitations*’ (after the abstract):

- “This study does not address quality of life”

4. *The authors should specifically focus on long-term outcomes (at least 6 months), for acute short-term outcomes (typically 6-8 weeks) are of very limited relevance to patients (and also of limited utility to clinicians).*

Similar to #3, this is also a valid suggestion. We acknowledge the reviewer’s consistent emphasis on the long-term outcomes of antidepressant trials (McPherson and Hengartner 2019; Hengartner 2020). Nevertheless, we here aim to describe the effects of different treatment strategies after nonresponse to an initial (set of) treatments, for which mostly short-term studies were designed. Therefore, for long term outcome data (e.g. as suggested after 6 months) we expect that this information will simply not be available. However, if these data turn out to be available, they definitely are of interest. (Deshauer et al. 2008; Hengartner 2020). Therefore, we added an analysis of long-term data to page 11 of the manuscript:

“We address long-term outcomes by additionally analyzing the primary outcomes at a treatment duration of 6 months or longer, if these data is available (Deshauer et al. 2008; Hengartner 2020). We

will exclude studies from the statistical synthesis if no primary endpoint data for the 4+ weeks period can be provided (Furukawa et al. 2016).”

5. *The definition of no treatment (<50% of patients receiving antidepressants) is problematic. According to this definition, patients that received a non-pharmacological treatment (psychological and/or psychosomatic, e.g. exercise) are considered untreated.*

We thank the reviewer for pointing this relevant unclarity in our definition. We meant to include any antidepressant treatment and not just antidepressant drugs. To clarify the definition, we amended the manuscript on page 10:

- “No treatment (NT; applies in case TAU involved virtually no intervention, defined as < 50% of patients receiving any antidepressant treatment (including pharmacotherapy, psychological therapy and/or neuromodulation treatment); patients know they will not receive active treatment after the trial)”

6. *How do the authors intend to account for varying study duration? Will they report short- and long-term results separately? This could be a serious issue, because I assume that trials testing a psychological intervention last longer, on average, than trials testing a pharmacological or neuromodulation intervention.*

We thank the reviewer for addressing this important issue. We acknowledge that this approach would favour studies with short term treatments, if we take endpoints for those short-term (e.g. psychopharmacological) studies and combine this with results after only half of (e.g. the psychotherapeutic) prolonged interventions have been performed. We propose to tackle this by limiting ourselves to the primary endpoints for each study, because that is chosen at time-points where such treatments are most likely to show differential benefit. Furthermore, we will report long-term outcomes separately, see #4.

“We will use the original author’s primary endpoint, ranging from 4 weeks or longer but less than 6 months, for analysis of the acute phase outcome data. We address long-term outcomes by additionally analyzing the primary outcomes at a treatment duration of 6 months or longer, if these data are available (Deshauer et al. 2008; Hengartner 2020). We will exclude studies from the statistical synthesis if no primary endpoint data for the 4+ weeks period can be provided (Furukawa et al. 2016).”

#### **Comments of dr. Jiaqiong Xu**

1. *This is a well-written manuscript. The statistical analysis plans are presented clearly and well. I don’t have any comments.*

We thank the reviewer for these kind words.

## References

- Berlim, M. T., and G. Turecki. 2007. 'What is the meaning of treatment resistant/refractory major depression (TRD)? A systematic review of current randomized trials', *Eur Neuropsychopharmacol*, 17: 696-707.
- Deshauer, Dorian, David Moher, Dean Fergusson, Ester Moher, Margaret Sampson, and Jeremy Grimshaw. 2008. 'Selective serotonin reuptake inhibitors for unipolar depression: a systematic review of classic long-term randomized controlled trials', *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, 178: 1293-301.
- Furukawa, T. A., G. Salanti, L. Z. Atkinson, S. Leucht, H. G. Ruhe, E. H. Turner, A. Chaimani, Y. Ogawa, N. Takeshima, Y. Hayasaka, H. Imai, K. Shinohara, A. Suganuma, N. Watanabe, S. Stockton, J. R. Geddes, and A. Cipriani. 2016. 'Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis', *BMJ Open*, 6: e010919.
- Hengartner, M. P. 2020. 'How effective are antidepressants for depression over the long term? A critical review of relapse prevention trials and the issue of withdrawal confounding', *Ther Adv Psychopharmacol*, 10: 2045125320921694.
- McPherson, Susan, and Michael P. Hengartner. 2019. 'Long-term outcomes of trials in the National Institute for Health and Care Excellence depression guideline', *BJPsych Open*, 5: e81-e81.
- Sivertsen, H., G. H. Bjørkløf, K. Engedal, G. Selbæk, and A. S. Helvik. 2015. 'Depression and Quality of Life in Older Persons: A Review', *Dement Geriatr Cogn Disord*, 40: 311-39.