

Letter to the Editor

Do not blame the SSRIs: blame the Hamilton Depression Rating Scale

Dear Editor,

In their recent paper entitled ‘Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis’ (1), Jakobsen et al. conclude that ‘The observed harmful effects seem to outweigh the potential small beneficial clinical effects of SSRIs, if they exist’ (1). In a follow-up article in a Danish popular science journal (Videnskab.dk), Jakobsen is quoted for the following statement on the selective serotonin reuptake inhibitors (SSRIs) (freely translated from Danish): ‘We are dealing with medicine that affects important neurotransmitters in the brain and has severe side effects. To justify giving this to people, we have to be sure that it works against depression. But it doesn’t’ (2). This message was broadly disseminated via the Danish news media in the days following the publication of the meta-analysis.

It is very important to communicate research findings to the general public. However, when making as blunt a statement as the one outlined above, which is likely to affect both the opinion and behaviour of individuals (for instance, the adherence to SSRI treatment or the likelihood of accepting SSRI treatment when indicated), researchers are ethically obliged to ensure that their interpretation of their results is completely unchallengeable. Below, I will make the argument that the interpretation made by Jakobsen et al. is far from unchallengeable, and that the statements in the paragraph above are therefore highly questionable.

Jakobsen et al. have performed a very extensive systematic search of both published and unpublished results of randomised clinical trials (RCTs) comparing the effect of SSRIs with that of placebo (1). They are to be complimented for that effort. Most of the RCTs included in the meta-analysis used the total score on the 17-item or the 21-item Hamilton Depression Rating Scale (HDRS) (3) as the outcome measure, and the primary results in the article by Jakobsen et al. (1) therefore also refer to the HDRS. The results of their meta-analyses ‘showed that SSRIs versus placebo significantly reduced the HDRS score (mean difference

–1.94 points; 95% CI –2.50 to –1.37; $p < 0.00001$)’ (1). Jakobsen et al. consider this ‘numerical’ superiority of the SSRIs over placebo on the HDRS to be below the threshold for clinical significance (3 points on the HDRS), which was suggested by the National Institute for Clinical Excellence (NICE) from the United Kingdom. Thus, the difference between 1.94 and 3 is essentially what leads Jakobsen et al. to the conclusion that there are only ‘small beneficial clinical effects of SSRIs, if they exist’ (1). The rationale for this conclusion may seem bulletproof, but there is a fundamental problem with the HDRS as outcome measure, which Jakobsen et al. do not take into consideration.

The HDRS was developed in 1960 (3) and consists of 17 symptom items in the original version (the 21-item version was never intended to be used for severity measurement in depression according to the scale’s developer Max Hamilton (4)). In the RCTs included in the meta-analysis by Jakobsen et al., the total score of these items is used as a measure for the overall severity of depression – and reduction in the HDRS total score over time is used as a measure of clinical improvement. In order for the total score to actually contain this clinical information, the HDRS must meet two fundamental criteria: (I) the total score of the items must correlate with evaluations of depressive severity made by clinical experts (gold standard), and (II) each of the items must convey unique information regarding the severity of the latent syndrome being measured, that is, depression (this is commonly referred to as ‘unidimensionality’ or ‘scalability’ (5)). In two landmark studies from 1975 (6) and 1981 (7), respectively, Bech et al. demonstrated that the original HDRS met none of these two criteria. This lack of psychometric validity of the HDRS has been confirmed in a large number of studies since then (8–15). Therefore, the total score of the HDRS cannot be considered as a clinically valid measure of the severity of depression (5,16). As the conclusions made by Jakobsen et al. (1) are based on the results of analyses of HDRS total scores, it entails that they are not clinically valid either.

Letter to the Editor

The landmark studies by Bech et al. (6,7) also demonstrated that although the total score of the HDRS is not a valid measure of depression severity, the scale contains a subscale of six items, that meets both of the validity criteria (clinical validity and unidimensionality/scalability) described in the paragraph above. These six items are as follows: item 1 – depressed mood; item 2 – guilt feelings; item 7 – work and interests; item 8 – psychomotor retardation; item 10 – psychic anxiety; and item 13 – somatic symptoms general (6). The subscale defined by these six items is now commonly referred to as ‘Hamilton-6’ (HDRS₆) (5). As opposed to the HDRS, the psychometric validity of the HDRS₆ has been confirmed numerous times (8–15) since its derivation from the HDRS in 1975 (6). Importantly, when using the total score on the HDRS₆ as outcome measure in RCTs of SSRIs (and related antidepressants) versus placebo, the effect sizes are markedly larger than those obtained when using the total score of the HDRS as outcome measure (17–20). There are two reasons for this difference, namely (i) the superior psychometric properties of HDRS₆ compared with HDRS, and (ii) the fact that three of the items in the HDRS (item 12 – somatic symptoms, gastrointestinal; item 14 – genital symptoms; and item 16 – loss of weight) tap into three common side effects of the SSRIs, namely diarrhoea/constipation, loss of libido, and loss of appetite. Indeed, in a recent meta-analysis of RCTs comparing SSRIs and placebo by Hieronymus et al. (18), these three HDRS items were the only ones yielding negative effect sizes – whereas the effect sizes of the remaining 14 items were positive. Thus, the HDRS contains an inherent bias against the SSRIs due to the side-effect profile of this class of drugs. Notably, none of these three problematic items are included in the HDRS₆. As the wanted effect (antidepressant effect in this context) and unwanted effects (side effects) are ideally evaluated independently in clinical studies (21–23), the HDRS₆ is an ideal measure of the wanted effects of antidepressant agents (21,24).

It should be mentioned that RCTs using either the Montgomery–Asberg Depression Rating Scale (MADRS) (25) or the Beck Depression Inventory (BDI) (26) as outcome measures were also included in the meta-analysis by Jakobsen et al. (1). The psychometric problems associated with the MADRS (27) and the BDI (6,28) are, however, equivalent to those mentioned in relation to the HDRS, so this makes little difference.

When confronted with the shortcomings of the HDRS, Jakobsen stated (freely translated from Danish): ‘We have used the conducted research as point of reference. You may have a fantasy that if a

different scale had been used, the result would have been different. But that is very theoretical’ (2). Using terms such as ‘fantasy’ and ‘theoretical’ in this context does not seem particularly fitting as there are published studies documenting that when using a psychometrically valid depression rating scale (HDRS₆) as outcome measure, the clinical superiority of SSRIs over placebo is quite consistent (17,18).

For the reasons outlined above, I strongly suggest that not only independent researchers like Jakobsen et al. (1), but also organisations like the NICE, the pharmaceutical industry, and the pharmaceutical evaluation authorities, such as the US Food and Drug Administration and the European Medicines Agency, will no longer consider the total score on the HDRS (or the MADRS or BDI for that reason) as being a valid outcome measure in studies of antidepressants – because this practice is in conflict with the results of a very large body of literature based on clinical psychometric research. Furthermore, it is my hope that Jakobsen et al. (1) will see this comment on their work as an encouragement to reanalyse their data using the HDRS₆ total score as outcome measure. This would be a highly clinically relevant contribution to the literature.

Conflicts of Interest

The author declares no conflicts of interest.

Søren Dinesen Østergaard
Psychosis Research Unit, Department of Clinical
Medicine, Aarhus University, Risskov, Denmark
E-mail: soeoes@rm.dk

References

1. JAKOBSEN JC, KATAKAM KK, SCHOU A et al. Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. *BMC Psychiatry* 2017;**17**:58.
2. Videnskab.dk. 9 February 2017. Available at <http://videnskab.dk/krop-sundhed/nyt-studie-lykkepiller-har-ingen-effekt-paa-depression> (in Danish).
3. HAMILTON M. A rating scale for depression. *J Neurolog Neurosurg Psychiatry* 1960;**23**:56–62.
4. HAMILTON M. The Hamilton Depression Scales. In: Sartorius N, Ban TA, editors. *Assessment of Depression*. Berlin, Germany: Springer, 1986.
5. BECH P. *Clinical Psychometrics*. Oxford, UK: Wiley-Blackwell, 2012.
6. BECH P, GRAM LF, DEIN E, JACOBSEN O, VITGER J, BOLWIG TG. Quantitative rating of depressive states. *Acta Psychiatr Scand* 1975;**51**:161–170.

7. BECH P, ALLERUP P, GRAM LF et al. The Hamilton Depression Scale. Evaluation of objectivity using logistic models. *Acta Psychiatr Scand* 1981;**63**:290–299.
8. OSTERGAARD SD, BECH P, TRIVEDI MH, WISNIEWSKI SR, RUSH AJ, FAVA M. Brief, unidimensional melancholia rating scales are highly sensitive to the effect of citalopram and may have biological validity: implications for the Research Domain Criteria (RDoC). *J Affect Disord* 2014;**163**:18–24.
9. OSTERGAARD SD, MEYERS BS, FLINT AJ et al. Measuring psychotic depression. *Acta Psychiatr Scand* 2014;**129**:211–220.
10. OSTERGAARD SD, BECH P, MISKOWIAK KW. Fewer study participants needed to demonstrate superior antidepressant efficacy when using the Hamilton Melancholia Subscale (HAM-D(6)) as outcome measure. *J Affect Disord* 2016;**190**:842–845.
11. LICHT RW, QVITZAU S, ALLERUP P, BECH P. Validation of the Bech-Rafaelsen Melancholia Scale and the Hamilton Depression Scale in patients with major depression; is the total score a valid measure of illness severity? *Acta Psychiatr Scand* 2005;**111**:144–149.
12. BECH P, FAVA M, TRIVEDI MH, WISNIEWSKI SR, RUSH AJ. Factor structure and dimensionality of the two depression scales in STAR*D using level 1 datasets. *J Affect Disord* 2011;**132**:396–400.
13. MARTINY K, REFSGAARD E, LUND V et al. The day-to-day acute effect of wake therapy in patients with major depression using the HAM-D6 as primary outcome measure: results from a randomised controlled trial. *PLoS One* 2013;**8**:e67264.
14. KYLE PR, LEMMING OM, TIMMERBY N, SONDERGAARD S, ANDREASSON K, BECH P. The validity of the different versions of the Hamilton Depression Scale in separating remission rates of placebo and antidepressants in clinical trials of major depression. *J Clin Psychopharmacol* 2016;**36**:453–456.
15. KORNER A, LAURITZEN L, ABELSKOV K et al. Rating scales for depression in the elderly: external and internal validity. *J Clin Psychiatry* 2007;**68**:384–389.
16. BAGBY RM, RYDER AG, SCHULLER DR, MARSHALL MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry* 2004;**161**:2163–2177.
17. BECH P, CIALDELLA P, HAUGH MC et al. Meta-analysis of randomised controlled trials of fluoxetine v. placebo and tricyclic antidepressants in the short-term treatment of major depression. *Br J Psychiatry* 2000;**176**:421–428.
18. HIERONYMUS F, EMILSSON JF, NILSSON S, ERIKSSON E. Consistent superiority of selective serotonin reuptake inhibitors over placebo in reducing depressed mood in patients with major depression. *Mol Psychiatry* 2016;**21**:523–530.
19. BECH P, KAJDASZ DK, PORSDAL V. Dose-response relationship of duloxetine in placebo-controlled clinical trials in patients with major depressive disorder. *Psychopharmacology (Berl)* 2006;**188**:273–280.
20. BECH P, BOYER P, GERMAIN JM et al. HAM-D17 and HAM-D6 sensitivity to change in relation to desvenlafaxine dose and baseline depression severity in major depressive disorder. *Pharmacopsychiatry* 2010;**43**:271–276.
21. BECH P. Applied psychometrics in clinical psychiatry: the pharmacopsychometric triangle. *Acta Psychiatr Scand* 2009;**120**:400–409.
22. BECH P, FAVA M, TRIVEDI MH, WISNIEWSKI SR, RUSH AJ. Outcomes on the pharmacopsychometric triangle in bupropion-SR vs. buspirone augmentation of citalopram in the STAR*D trial. *Acta Psychiatr Scand* 2012;**125**:342–348.
23. BECH P, GEFKE M, LUNDE M, LAURITZEN L, MARTINY K. The pharmacopsychometric triangle to illustrate the effectiveness of T-PEMF concomitant with antidepressants in treatment resistant patients: a double-blind, randomised, sham-controlled trial revisited with focus on the patient-reported outcomes. *Depress Res Treat* 2011;**2011**:806298.
24. PAPAPOSTAS GI, OSTERGAARD SD, IOVIENO N. The nature of placebo response in clinical studies of major depressive disorder. *J Clin Psychiatry* 2015;**76**:456–466.
25. MONTGOMERY SA, ASBERG M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;**134**:382–389.
26. BECK AT, WARD CH, MENDELSON M, MOCK J, ERBAUGH J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;**4**:561–571.
27. BECH P, ALLERUP P, LARSEN ER, CSILLAG C, LICHT RW. The Hamilton Depression Scale (HAM-D) and the Montgomery-Asberg Depression Scale (MADRS). A psychometric re-analysis of the European genome-based therapeutic drugs for depression study using Rasch analysis. *Psychiatry Res* 2014;**217**:226–232.
28. BOUMAN TK, KOK AR. Homogeneity of Beck's Depression Inventory (BDI): applying Rasch analysis in conceptual exploration. *Acta Psychiatr Scand* 1987;**76**:568–573.