What is the threshold for a clinically relevant effect? The case of Major Depressive Disorders

Article *in* Depression and Anxiety · May 2014

DOI: 10.1002/da.22249 · Source: PubMed

5 authors, including:

Pim Cuijpers
VU University Amsterdam
704 PUBLICATIONS **22,176** CITATIONS
SEE PROFILE

Erick H Turner
Oregon Health and Science University
93 PUBLICATIONS **3,577** CITATIONS
SEE PROFILE

Sander L Koole
VU University Amsterdam
135 PUBLICATIONS **4,407** CITATIONS
SEE PROFILE

Depression AND Anxiety The official journal of ADAA

What is the threshold for a clinically relevant effect? The case of major depressive disorders

Journal:	*Depression and Anxiety*
Manuscript ID:	DA-13-481.R1
Wiley - Manuscript type:	Theoretical Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Cuijpers, Pim; VU University Amsterdam, Psychology and Education Turner, Erick; Portland Veterans Affairs Medical Center, Behavioral Health and Neurosciences Division Koole, Sander Van DIjke, Annemiek Smit, Filip
Keywords:	Antidepressants, Depression, dysthymic disorder, mood disorders, Problem Solving Therapy, Pharmacotherapy, IPT/interpersonal psychotherapy, CBT/cognitive behavior therapy

What is the threshold for a clinically relevant effect? The case of major depressive disorders

Pim Cuijpers, Ph.D. [1,2,3,§]

Erick H. Turner, M.D. [4,5]

Sander L. Koole, Ph.D. [1,2]

Annemiek van Dijke, Ph.D. [6]

Filip Smit, Ph.D. [1,2,7,8]

[1] Department of Clinical Psychology, VU University Amsterdam, The Netherlands

[2] EMGO Institute for Health and Care Research, VU University and VU University Medical Center Amsterdam, The Netherlands

[3] Leuphana University, Lünebrug, Germany

[4] Behavioral Health and Neurosciences Division, Portland Veterans Affairs Medical Center, Portland, OR, USA

[5] Departments of Psychiatry and Pharmacology, Oregon Health and Science University, Portland, OR, USA

[6] Delta Psychiatrisch Centrum, Poortugaal, The Netherlands

[7] Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, The Netherlands

[8] Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands

[§] Corresponding author: Pim Cuijpers, Ph.D., Professor of Clinical Psychology Department of Clinical Psychology and EMGO Institute for Health and Care Research, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands, tel. # 31 20 5988757, fax. # 31 20 5988758, email: p.cuijpers@vu.nl.

Word count: 1,836

Abstract

Background: Randomized trials can show whether a treatment effect is statistically significant and can describe the size of the effect. There are, however, no validated methods available for establishing the clinical relevance of these outcomes. Recently, it was proposed that a standardised mean difference of 0.50 be used as cut-off for clinical relevance in the treatment of depression.

Methods: We explore what the effect size means and why the size of an effect has little bearing on its clinical relevance. We will also examine how the "minimally important difference", as seen from the patient perspective, may be helpful in deciding where the cut-off for clinical relevance should be placed for a given condition.

Results: Effect sizes in itself cannot give an indication of the clinical relevance of an intervention, because the outcome itself determines the clinical relevance and not only the size of the effects. The "minimal important difference" could be used as a starting point for pinpointing the cut-off for clinical relevance. A first, rough attempt to implement this approach for depression resulted in a tentative clinical relevance cut-off of SMD=0.24. Using this cut-off, psychotherapy, pharmacotherapy and combined treatment have effect sizes above this cut-off.

Discussion: Statistical outcomes can not be equated with clinical relevance. The "minimal important difference" may be used for pinpointing the cut-off for clinical relevance, but more work in this area is needed.

Keywords: depression; effect size; clinical relevance; minimal important difference.

Introduction

When does a treatment of depression have a clinically relevant effect? Despite hundreds of randomized controlled trials on antidepressant medications, psychotherapies and several other types of treatment, we still have no proper way of determining whether a treatment's effects are clinically relevant. In 2004, this issue was placed in the limelight of scientific attention when the National Institute of Clinical Excellence (NICE) proposed that a standardised mean difference (SMD or Cohen's d) of 0.50 represents a clinically relevant effect.[1] ~~By implication, t~~Treatments with smaller effects were deemed to be clinically irrelevant. This resulted in several highly-cited studies announcing that antidepressant medication does not have a clinically relevant effect in mild to moderate depression, because the SMD was below the cut-off of SMD=0.50.[2,3] NICE abandoned this position in its 2009 guideline,[4] but this ~~was~~ went apparently unnoticed, ~~as~~ given that the 2004 cut-off has received considerable attention in the media and scholarly journal articles. Thus many popular and scientific reports continue to state that antidepressants should only be used in severe depression, where the SMD exceeds 0.50.

In this paper we will show why the cut-off at SMD=0.50 is inappropriate for determining clinical relevance. We further propose an alternative method that explicitly considers the patient's perspective in deciding whether treatments are clinically relevant. Finally, we ~~show~~ illustrate how this alternative method can be applied to determine a cut-off for clinical relevance in the treatment of depression, and suggest that this method that can be used to develop cut-offs for other disorders as well.

Problems with using a fixed value of SMD as indicator of clinical relevance

When the concept of SMD was developed in the 1970s, the main advantage was that it gave an indication of the *size* of an effect, rather than whether the effect was significantly different from zero.[5] The latter is not very informative, because a statistical test depends on the sample size, effect size, and its variance. Thus large studies, compared to small studies, are more likely to find statistically significant effects. In contrast to the p-value, the SMD captures the size of an effect, regardless of its significance.

The SMD describes the difference between the treated and control groups in terms of standard units (standard deviations). So, a SMD of 1 means that the treated and the control group differ by one standard deviation from each other at post-test.

But how can one determine whether such a SMD is clinically relevant? Jacob Cohen, a pioneer of the concept of the SMD, proposed values of 0.2, 0.5, and 0.8 as small, medium, and large, respectively.,[6] ~~while~~ It should be noted that Cohen acknowledged~~ing~~ he lacked empirical data to support these delineations. Despite this acknowledgment, NICE, without providing scientific justification, later adopted the "medium" value as its cut-off for clinical relevance.

An empirical method of defining cut-offs for SMDs was developed by Lipsey and Wilson.[7] They reviewed all meta-analyses that were available by that time in the fields of psychology, behavioural and educational sciences, and computed the median effect size of all those meta-analyses. This appeared to be SMD=0.45, not very different from the SMD=0.50 suggested by Cohen as a medium effect. So, if we would have to choose an effect size of medium size, the SMD=0.45 would have been an appropriate candidate. Taking this approach, the next step ~~we~~ would be ~~have~~ to update our database of meta-analyses with the many studies that have taken place since 1990.; ~~we~~ Research would then proceed by calculating SMDs for different disciplines, as effect sizes in one discipline may not be comparable ~~with~~ to those in others (see below); and ~~we~~ it would be needed to account for the possibility that many of the effect sizes have been inflated by publication bias.

However, meta-analytic approaches cannot overcome more a fundamental ~~interpretational~~ problem, that being how to translate ~~in relating~~ the SMD into clinical relevance. The SMD is a purely statistical construct, with no reference to the domain-specific meaning of what is being assessed. ~~Therefore~~In reality, the meaning of the SMD varies ~~strongly~~ considerably for different subdomains and outcomes. As indicated, the SMD captures the difference between two groups in standard units, which still has little bearing on the clinical relevance of the effect. To illustrate, a SMD=0.1 in terms of years of survival ~~will~~ would be considered by most clinicians as a very important and strong effect, ~~while~~ whereas the same SMD of 0.1 in terms of more "social skills" or "knowledge about depression" would likely ~~be~~ not be considered clinically meaningful by most. ~~So~~Thus, there is little correspondence~~no correlation~~ between the size of SMD

and its clinical relevance. Therefore, proclaiming that an SMD of 0.50 is the cut-off for a clinically relevant effect in the treatment of depression is inaccurate~~dequate~~ and misleading.

There have been several previous attempts to define what a clinical relevant outcome is,[8] including "clinical significant change" proposed by Jacobson and Truax.[9] According to these methods, a threshold is defined above which it is not probable that change has happened by chance. All of these methods, however, are based on statistical procedures. Statistical procedures, however, cannot capture the domain-specific factors that determine whether a change is meaningful in a particular clinical context. Consequently, statistical procedures alone are inadequate for determining clinical relevance.

Formatted: Superscript

Formatted: Superscript

An alternative method of defining clinical relevance in depression

~~Perhaps a better~~A more valid way of determining the clinical relevance of an effect ~~is~~ may be to use the "minimal important difference" (MID) from the patient perspective, because ultimately patients should be the judges of their own health.[10~~8~~] The MID can be defined as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management".[11~~9~~]

In a recent study,[12~~0~~] the MID was estimated using an abbreviated version of the SF-36,[13~~1~~] the SF-6D. The SF-6D was used to obtain preference-based measures of health, called "utilities" (U), ~~that~~ with the two ends ~~are~~ anchored at 0 (death) and 1 (perfect health).[12~~0~~] In a next step, Brazier and colleagues used the data from 11 studies in different patient groups,[12~~0~~] to ascertain whether the patients had experienced a MID at one year follow-up compared to baseline. ~~They asked t~~The patients ~~how they would~~ rated their health in general at that~~is~~ moment compared to one year ago. If a patient indicated that health was "somewhat better than 1 year ago" or "somewhat worse", the authors assumed that a MID had occurred. ~~Then~~ U~~u~~sing the SF-6D based utilities, U, the authors ~~could~~ then calculated how large the MID was on the scale of U. It was then observed that, on average, the MID ~~was~~ corresponded~~ing~~ with U=0.04 or larger.

The MID estimate is, however, not sufficient for deciding ~~what~~ the cut-off for a clinically relevant effect size ~~is such as~~ in terms of the SMD. To find this~~at~~ cut-off, we

have to ~~map~~ transform the MID of U=0.04 ~~o~~into the SMD; so, how large is the SMD when U=0.04~~.~~? Another recent study,[1~~42~~] used vignettes to estimate how much change in health states (again on a utility scale with 0 for death and 1 for optimal health) corresponds with an effect size (SMD). General practitioners who had been trained in the recognition and management of mental disorders, rated vignettes with varying severity of several mental disorders~~,~~ including depression. Because an SMD of 1.0 indicates a difference of 1 standard deviation (SD) between two groups, the researchers~~y~~ defined variation in disorder severity in SD units. The mental health summary scale of the SF-12 was used to anchor the vignettes in SD decrements, and the vignettes were created based on the criteria for major depression, remission (meeting ICD-10 diagnostic criteria at some time in the past 12 months but not currently), and the symptom profiles of depression according to the ICD-10. This approach resulted in an estimate of an increase of the health state (on the scale of 0 to 1) of U=0.17 for every standard deviation of improvement. Thus, according to this estimate, a SMD=1 for depression treatment corresponds with an increase in health of U=0.17. If we then take the MID of U=0.04 ~~we~~ that Brazier and colleagues found earlier, ~~we~~ one could say that the MID for the treatment of depression is (0.04/0.17=) SMD=0.24.

For example, suppose we monitor treatment response with the CES-D in a group of depressed patients and the standard deviation of the CES-D in this group is 9. One patient has a score of 32 at baseline and 30 after treatment. Does this patient meet criteria for MID? We can compute: SMD=(32-30)/9=0.22. This is lower than the agreed MID threshold of 0.24 and we are therefore not satisfied that this patient is changed at a level that can be subjectively appreciated by this patient. However, patients who changed 3 points (or more) on the CES-D would indicate their change (3/9=0.33) as 'minimally important' in the sense that their improvements are subjectively detectable.

The foregoing is of course only a very rough estimate of the cut-off for clinical relevance, and only illustrates how such cut-offs may be estimated. It should be clear, however, that much more research is needed to establish such cut-offs more precisely.

Are current treatments of depression clinically relevant?

~~But s~~Suppose that the cut-off ~~would~~ is indeed ~~be~~ SMD=0.24~~,~~. Could we then conclude that current treatments for depression are clinically relevant? This paper is

only meant to illustrate the method, and it is beyond the scope of this paper to conduct a full search for meta-analyses of pharmacotherapy and psychotherapy. However, to give an overall first impression, we selected meta-analyses with the same comparator (pill placebo) and adjusted for publication bias. Table 1 displays the effect sizes for the most important treatment of depression, when they are compared with pill placebo control groups. The effect size of pharmacotherapy was taken from the study from Turner and colleagues,[15~~3~~] in which the effect size was ~~adjusted~~ corrected for publication bias. The effect size for psychotherapy was taken from a recent meta-analysis of studies comparing psychotherapy with pill placebo.[16~~4~~]

The effect size for combined psychotherapy and pharmacotherapy has not been published before. ~~but~~ However, this estimate is based on studies which we selected from an existing database of studies on psychotherapy for adult depression,[17~~5~~] using the same methodology as in the meta-analysis of studies on psychotherapy versus pill placebo control groups. Because most types of antidepressant medications,[18~~6~~] as well as most types of psychotherapies,[19~~7~~,~~18~~-20] have comparable effect sizes (and because for each type of psychotherapy only a small number of placebo-controlled trials is available), we report overall effect sizes for the categories of psychotherapy and pharmacotherapy, without specifying effect sizes for each type of therapy and medication. In Table 1, we have also given the increase in utilities for each treatment based on the described estimate that a SMD of 1.0 corresponds with an increase of in health-related quality of life of U=0.17.

As can be seen, all treatments have effect sizes ~~that are higher than~~above the cut-off of d=0.24. Especially the combination of pharmacotherapy and psychotherapy scores well above the cut-off. There is a caveat for psychotherapy, however. While the point estimate of its effect size (d=0.25) is just above the threshold for clinically relevance, its 95% confidence interval~~s~~ straddles it, so further studies are needed to confirm whether the effect truly lies above the threshold.[21~~19~~] ~~All this~~These specific numbers should be considered with caution, given the many uncertainties in calculating the cut-off and the effect sizes of the different treatments.

Discussion

Currently, there is no ~~proper~~ valid threshold for clinical relevance that can be expressed on a SMD scale of standard units. ~~,~~ Despite its popularity, the notion ~~and that the suggestion~~ that an effect size of d=0.50 can be considered as cut-off for clinical relevance is neither empirically nor theoretically supported. Moving beyond the SMD~~Instead~~, we proposed the "minimal important difference" as a starting point for pinpointing the cut-off for clinical relevance. Future research may elaborate on this approach. A first, rough attempt to implement this approach for depression resulted in a tentative clinical relevance cut-off of SMD=0.24. In the study in which an estimate of the MID was made, no depressed samples were used, and this may have influenced the outcomes. In that same study the MID was also estimated with another instrument, the EQ-5D, which resulted in a different, higher estimate of the MID. We chose to base our estimate on the SF-6D not only because it yielded a lower, more conservative estimate, but also because SF scales were used to transform utilities into effect sizes. Nevertheless, this indicates that we are dealing with unstable values, which may result in thresholds higher or lower than ours. This proposed value should, therefore, considered only to be a preliminary estimate.

Furthermore, the MID itself may also depend on the circumstances and the larger context. For example, a change of one degree of body temperate may be less important in an adult with fever than in a child with fever, and one degree of change becomes important when the fever in the child increases. It is important therefore, that domain-specific MIDs are developed, which may serve as 'bench marks' for clinically relevant improvements in specific domains.

These caveats aside, ~~this~~ the present approach ~~may~~ provides direction as to how clinical relevance cut-offs may be ~~identified~~ determined for health problems such as depression, while avoiding an overly naïve way of ~~equating~~ confounding statistical ~~outcomes~~ effect sizes with clinical relevance.

Summary box

- Effect sizes is a purely are statistical concept; by itself, s that have little or no bearing it does not provide information on the clinical relevance of treatment effects.
- A standardized mean difference (SMD or Cohen's d) of 0.50 is not a valid threshold for clinically relevant effects of treatments for depressive disorders.
- Based on studies estimating the "minimal important difference" between health states from the patient perspective, and modelling studies to transform effect sizes to health states, it is possible to estimate a threshold for clinical relevance of treatments.
- Using this method we found that an effect size of SMD=0.24 may be consideredas a preliminary estimate of a cut-off for clinical relevance for the treatment of depression.

Contributors

The idea for this paper is the result of discussion among the authors about this subject. PC wrote the first version of this paper and all co-authors read the different versions of the paper critically and all contributed important content. PC is the nominated guarantor of this article.

Conflicts of Interest

None

Acknowledgements

None.

References

1. National Institute for Clinical Excellence. Depression: management of depression in primary and secondary care. Clinical practice guideline No 23. London: National Institute for Clinical Excellence, 2004.
2. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: A meta-analysis of data submiteed to the food and drug administration. PLoS Med 2008; 5: 250-68.
3. Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC et al. Antidepressant drug effects and depression severity: A patient-level meta-analysis. JAMA 2010; 303: 47-53.
4. National Institute for Clinical Excellence. Depression: The treatment and management of depression in adults. Partial update of clinical practice guideline No 23. London: National Institute for Clinical Excellence, 2009.
5. Cohen J. The earth is round (p<.05). Am Psychologist 1994; 49: 997-1003
6. Cohen J. Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum, 1988.
7. Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment confirmation from meta-analysis. Am Psychol 1993; 48: 1181-209.
8. Jacobsen N, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991; 59: 12-9.
9. Bauer S, Lambert MJ, Nielsen SL. Clinical significance methods: A comparison of statistical techniques. J Person Assessm 2004; 82: 60–70.
10. Stiggelbout AM, Van der Weijden T, De Wit MPT, Frosch D, Légaré F, Montori VM, et al. Shared decision making: really putting patients at the centre of healthcare. BMJ 2012; 344: e256 doi: 10.1136/bmj.e256
11. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Contr Clin Trials 1989; 10: 407–15.
12. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res 2005; 14: 1523–32.

13. Ware JE Jr, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey Manual and Interpretation Guide. Boston, MA: The Health Institute, New England Medical Centre, 1993.

14. Sanderson K, Andrews G, Corry J, Lapsley H. Using the effect size to model change in preference values from descriptive health status. Qual Life Res 2004; 13: 1255–64.

15. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008; 358: 252-60.

16. Cuijpers P, Turner EH, Mohr DC, Hofmann SG, Andersson G, Berking M, et al. Comparison of psychotherapies for adult depression to pill placebo control groups: A meta-analysis. Psychological Medicine 2013, epub ahead of print.

17. Cuijpers P, van Straten A, Warmerdam L, Andersson G. Psychological treatment of depression: A meta-analytic database of randomized studies. BMC Psychiatry 2008; 8(1): 36.

18. Gartlehner G, Hansen RA, Morgan LC, Thaler K, Lux L, Van Noord M, et al. Comparative Benefits and Harms of Second-Generation Antidepressants for Treating Major Depressive Disorder; An updated meta-analysis. Ann Intern Med 2011; 155: 772-85.

19. Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. J Consult Clin Psychol 2008; 76: 909-22.

2018. Barth J, Munder T, Gerger H, Nuesch E, Trelle S, Znoj H, et al. Comparative efficacy of seven psychotherapeutic interventions for depressed patients: A network meta-analysis. Plos Med 2013; 10: e1001454.

2119. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. Biol Psychiatry 2006; 59: 990-6.

Table 1. Standardized mean differences (SMDs) for current treatments of depression compared with pill placebo

	SMD	*95% CI*	*U* [a)]
Pharmacotherapy [b)]	0.31	0.27 ~ 0.35	0.05
Psychotherapy [c)]	0.25	0.14 ~ 0.36	0.04
Combined therapy [d)]	0.46	0.21 ~ 0.70	0.08

[a)] Increase of the health state (on the scale of 0 to 1)
[b)] Based on Turner et al., New Engl J Med 2008;[153] corrected for publication bias
[c)] Based on Cuijpers et al., 2013[164]
[d)] Data from our database of trials on psychotherapy for adult depression (www.evidencebasedpsychotherapies.org), only combined versus placebo (N=6)