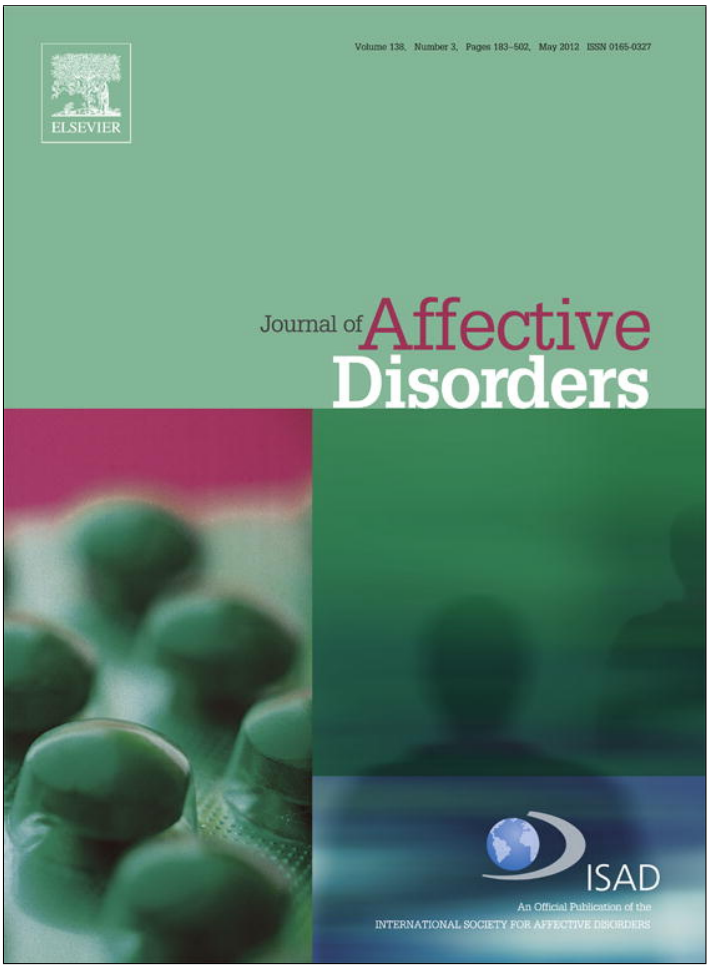

Journal of Affective Disorders 138 (2012) 183–191

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Review

Can effects of antidepressants in patients with mild depression be considered as clinically significant?

Ulrich Hegerl [a,*], Antje-Kathrin Allgaier [b], Verena Henkel [c], Roland Mergl [a]

[a] Department of Psychiatry, University of Leipzig, Semmelweisstr. 10, D-04103 Leipzig, Germany
[b] Department of Child and Adolescent Psychiatry, Ludwig-Maximilians-University (LMU) München, Nußbaumstraße 5a, D-80336 München, Germany
[c] Department of Psychiatry, Ludwig-Maximilians-University (LMU) München, Nußbaumstr. 7, D-80336 München, Germany

ARTICLE INFO

Article history:
Received 2 February 2011
Received in revised form 3 May 2011
Accepted 9 May 2011
Available online 8 June 2011

Keywords:
Antidepressants
Clinical trial
Depression
Placebo–verum difference
Hamilton Depression Rating Scale

ABSTRACT

Background: How to define clinical significance of antidepressants has become a matter of far-reaching clinical and regulatory consequences. A mean difference of at least 3 points on the Hamilton Depression Rating Scale (HAMD-17) between active treatment and placebo has been proposed as cut-off score for clinical significance in antidepressant trials.
Objective: We aimed to present arguments that this, and other commonly used related approaches to establish clinical significance are likely to be misleading and risky depriving patients with mild depression of efficient treatments.
Methods: These problems are exemplified with the data from a randomized placebo-controlled five-arm clinical trial with primary care patients with milder depressive syndromes (MIND-study).
Results and conclusions: Designs for studying clinical significance have to be distinguished from those assessing efficacy. Moreover, evaluation of the clinical significance of psychotherapy as a possible alternative to antidepressants faces the problem of how to define a valid control group where blinding of neither therapists nor patients is possible.

Contents

* Corresponding author. Tel.: +49 341 97 24530; fax: +49 341 97 24539.
E-mail address: Ulrich.Hegerl@medizin.uni-leipzig.de (U. Hegerl).

0165-0327/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jad.2011.05.015

1. Introduction

Mild depression is highly prevalent in non-psychiatric populations, especially primary care (28.5%) (Vuorilehto et al., 2005). However, there are considerable uncertainties and contradictions in guideline recommendations for these patients. For example, "watchful waiting" or "active monitoring" is considered a reasonable option according to past NICE guidelines (Middleton et al., 2005; NICE, 2004) or the national guidelines for care of unipolar depressive disorders in Germany (DGPPN et al., 2009) whereas active treatment with antidepressants is recommended by the APA(2000b) (see also Bauer et al., 2007). Revised NICE guidelines (NICE, 2009) still recommend "active monitoring" and offer new specifics (page 111):

> "For people who, in the judgement of the practitioner, may recover with no formal intervention, or people with mild depression who do not want an intervention, or people with subthreshold depressive symptoms who request an intervention:

- discuss the presenting problem(s) and any concerns that the person may have about them;
- provide information about the nature and course of depression;
- arrange a further assessment, normally within 2 weeks;
- make contact if the person does not attend follow-up appointments."

In the case of patients with mild depression who seek intervention, the NICE guidelines advise physicians to use antidepressants or CBT only if low-intensity treatments (like guided self-help based on CBT, computerized CBT, and group exercise as well as sleep hygiene education) have been tried and found not effective (NICE, 2009). If patients with mild depression have a history of episodes with more severe intensity, antidepressants can be prescribed as first-line treatment (see also Davidson, 2010).

The importance of this topic stems from the large number of patients suffering from mild depression and the health economic consequences of treatment recommendations and regulations in this area (Cuijpers et al., 2007). For the development of guidelines and for decision makers in health care systems who have to allocate limited resources, a central question is whether the effects obtainable with antidepressants can be considered large enough to be clinically significant. The discussion is enriched by concerns that mild depressive mood swings as part of daily and sometimes bitter life are "psychiatrisized", possibly in line with the interest and intention of the pharmaceutical industry or other interest groups such as the psychotherapists or psychiatrists to "create" new customers by lowering the disease threshold.

The aim of this article is

1. to explain why the term "mild depression" is misleading;
2. to draw attention to the fact that the present procedure to measure clinical significance can prove grossly misleading and
3. to compare psychotherapy as an alternative to antidepressant pharmacotherapy concerning the evidence of clinically significant effects.

2. The term "mild depression" is misleading

"Mild depression" is a problematic term for several reasons. One is that there is no consensus how to measure it. Mild depression can be defined by the presence of a certain number of diagnostic criteria. This approach is used in psychiatric classification systems like ICD-10 (WHO, 1992) or DSM-IV-TR (APA, 2000a). Another approach is to define mild depression by a certain range of sum rating scores using rating scales such as the Hamilton Depression Rating Scale (HAMD) (Hamilton, 1960). There is only partial overlap between these two approaches (Klinkman et al., 1997).

Within these two approaches, different definitions are used. According to the American Psychiatric Association Task Force for the Handbook of Psychiatric Measures (Task Force for the APA Handbook of Psychiatric Measures, 2000) mild depression is characterized by a sum score of 8 to 13 points in the 17-item version of the HAMD. This definition is used in a recent meta-analysis of antidepressant drug effects and depression severity (Fournier et al., 2010). Other meta-analyses, however, use markedly different definitions of mild depression (Angst et al. (1993): HAMD-17 sum scores $\leq$21 points; Shelton et al. (2007): HAMD-17 sum scores $\leq$19 points; de Maat et al. (2007): HAMD-17 interval between 12 and 19.9 points).

Aside from inconsistencies in these definitions the term "mild depression" suggests a rather benign condition. Yet, even subthreshold and minor depressive disorders, not even meeting the criteria necessary for diagnosing mild depression, can prove serious disorders and can be associated with considerable deficits in psychological well-being (Nierenberg et al., 2010), a strong negative impact on life quality (Nierenberg et al., 2010), functional impairment (Rapaport and Judd, 1998), increased mortality as compared to non-depressed subjects (Cuijpers and Smit, 2002) as well as increased risks of suicidality (Angst and Merikangas, 1997; Fergusson et al., 2005) and transition to major depression (Fogel et al., 2006; Judd et al., 1998; Lyness et al., 2006). On these bases, mild depression is milder than severe depression, but not a mild disorder.

3. Do antidepressants have 'clinically significant' effects?

It becomes important to examine whether antidepressant treatment effects in mild depression are not only statistically significant, but large enough to be also clinically significant. Whereas the statistical significance of treatment effects is not questioned, the clinical significance has been challenged by recent meta-analyses (Fournier et al., 2010; Kirsch et al., 2008) which suggest that the amount of benefit of antidepressants in comparison with placebo escalates with intensity of depression symptoms and may be minimal in the case

of patients suffering from mild or moderate depression. Fountoulakis and Möller (2010) seriously challenged this view in a re-analysis of the data presented by Kirsch et al. (2008). Important flaws were detected for the calculations. Presentation of the results was shown to be selective whereas conclusions appeared to be overemphasized. Fountoulakis and Möller (2010) concluded from their re-analysis of the Kirsch data that efficacy of antidepressants was always existent and not restricted to a special degree of depression severity.

The above-mentioned reviews as well as national guidelines like the German guidelines for the treatment of unipolar depression (DGPPN et al., 2009) appear to adopt criteria for clinical significance proposed by NICE (2004). The NICE selected as one measure a drug–placebo difference of at least three points concerning the improvement compared to baseline in the HAMD-17 (Hamilton, 1960) sum score (ΔHAMD-17) as a threshold for clinical significance (NICE, 2004).

Having in mind the far reaching consequences of the definition of clinical significance for the care of depressed patients as well as for health political decisions, it is surprising that the method how to measure it is not more broadly discussed. Only a few articles brought this issue into focus in the context of depression (Bech, 2006). In contrast to the definition of clinical significance presented by NICE (2004), Bech recommends the use of standardized effect size statistics for the six-item version of the HAMD (HAMD-6) (Beck, 1967) to determine clinically significant antidepressant effects (Bech, 2006; Faries et al., 2000). This measure has the advantages of being unit-free and independent of sample size (in contrast to a p value) (Faries et al., 2000).

To declare that a particular treatment has no clinically significant effects implies that in daily practice provision of the treatment and alternatives such as active monitoring do not really make a relevant difference for the individual patient. When making such a statement, care should be taken not to underestimate treatment effects and to discard erroneously highly effective treatments. The situation is different than in efficacy studies. The aim of these studies is to prove in principle that the treatment has antidepressant effects. This is done by testing the outcome against placebo within a randomized controlled trial (RCT). Here, care has to be taken not to infer efficacy erroneously. Since the aims behind demonstrating efficacy and demonstrating clinical significance are different, study design and methodology have to be different. As we will see, there are compelling arguments that accepted procedure to deduce clinical significance from differences between active treatment and placebo in improvement in HAMD-17 scores, response rates or other parameters obtained within randomized clinical trials (RCTs) is likely to prove grossly misleading and risky depriving patients of efficient treatments.

3.1. Clinical significance of antidepressant treatment effects cannot be derived from intent-to-treat and last-observation-carried forward approaches

In RCT efficacy studies, intent-to-treat (ITT) and last-observation-carried-forward (LOCF) approaches are used. This conservative approach makes sense in order to avoid false positive results. However, such an analysis does not make sense when clinical significance is studied. The possible benefit of a treatment would be underestimated if we included in the analysis also patients who stopped treatment prematurely and were not treated properly for other reasons. The size of a treatment effect cannot be measured in patients who are not treated for whatever reason. Clinical significance should tell us how much the patient will improve when he takes and tolerates the drug and this can be best studied in the per protocol population. For the clinical decision to start or not to start a specific treatment in an individual patient many factors such as side effects, costs, treatment alternatives, patients' preferences etc. come into play, but balancing side effects and other costs and risks associated with antidepressant treatment effects is not the topic addressed by clinical significance. In the meta-analyses mentioned above (Fournier et al., 2010; Kirsch et al., 2008) the drug-placebo difference was measured based on ITT data with the LOCF approach. It has to be expected that the placebo–verum difference in ΔHAMD-17 would be clearly larger and possibly above three points also for mild depression when the per-protocol and not the ITT population would be analyzed. The differences between these two analyses can be considerable. According to a meta-analysis of major depressive disorder trials the drug–placebo success rate difference is 28.1% in the per-protocol population and only 18.8% in the ITT population (Stolk et al., 2003).

3.2. Differences between active treatment and placebo are underestimated because of adherence problems

In assessing clinical significance of treatment, it has to be assumed that the patients are adherent with their treatment. In the RCTs used to evaluate the clinical significance of antidepressants it has to be expected that even in the per-protocol population a large number of the patients take the drug irregularly or not at all. In RCTs with antidepressants, nonadherence, as measured mainly by the number of pills taken, has been found to be between 28 and 80% (Pampallona et al., 2002). This may be an underestimation of the problem because even when plasma levels are measured patients may take the drug the day before the visit and forget or not take it for other reasons the other days. As can be expected, the response rate is significantly higher in adherent than non-adherent patients (82.5% versus 55.8% with adherence being defined according to a composite index including questioning, serum levels and appointments kept). This difference is irrespective of the method applied to determine adherence (Akerblad et al., 2003). Moreover, it could be demonstrated that the number of depressed patients who reduced the dosage of the received antidepressant or transiently discontinued the drug was significantly lower in the group in which adherence was measured by serum levels of sertraline and desmethylsertraline, as compared to a control group (Akerblad et al., 2003). In addition, this study demonstrated the problems associated with adherence measurement by the finding that 37–70% of depressed patients showed treatment adherence, depending on the method applied.

In our own randomized controlled study of patients with milder forms of depression (MIND-study; (Hegerl et al., 2010) see below) a surprising finding was that 2 patients in the placebo arm had positive sertraline plasma levels. One probable explanation might be that these patients wanted to

be sure to receive the antidepressant, have thrown away the study drugs and have secured themselves sertraline via another physician or the internet. The consequence of the mentioned adherence problems is that even when analyzing the per-protocol population the active treatment/control difference would underestimate the real antidepressant effect a depressed patient could expect when taking the drug.

This is exemplified using the data from the MIND-study (Hegerl et al., 2010), a randomized, placebo-controlled, single-center, 10-week clinical trial involving primary care patients with milder forms of depression. Treatment arms were 1) sertraline (flexible dosages up to 200 mg/day) (n = 83), 2) placebo (n = 83), 3) manual-guided cognitive–behavioral group therapy (1 individual session and 9 group sessions per 90 min) (n = 61), 4) guided self-help group (control condition (n = 59)) and 5) treatment with sertraline or cognitive–behavioral group therapy according to patients' choice (n = 82).

The detailed study protocol was in accordance with the Declaration of Helsinki, as revised 1989, and Good Clinical Practice guidelines (International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2001). The study was approved by an independent Ethics Review Committee, and all subjects gave written informed consent.

When comparing the placebo–verum difference of improvement in HAMD-17 (ΔHAMD-17) between the ITT population with LOCF and the PP population excluding non-completers and those with non-adherence the placebo–verum difference was larger in the latter group (2.99 versus 2.07; see also Fig. 1).

3.3. Overrepresentation of non-responders in RCTs leads to an underestimation of the clinical significance in daily practice

The decision whether to treat a patient with mild depression is especially difficult in patients without previous treatment.

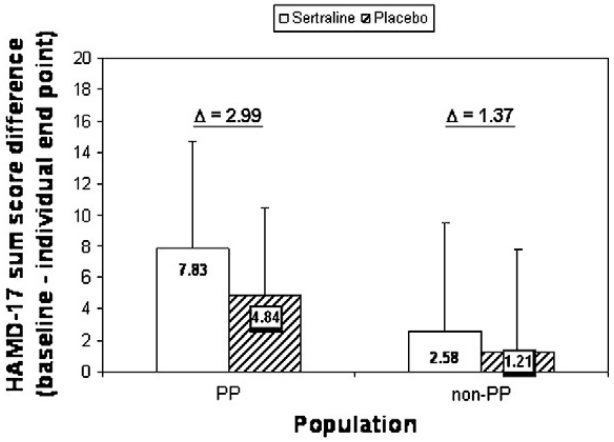

Fig. 1. Anti-depressive treatment effects and compliance in primary care patients with mild-to-moderate depressive disorders, based on data from the MIND study (Hegerl et al., 2010): Sertraline–placebo differences regarding HAMD-17 sum score changes (baseline — individual endpoint) in patients with low (non-PP) versus high (PP) compliance (according to ITT-LOCF analysis). Notes: HAMD-17: Hamilton Depression Rating Scale — 17-item version (Hamilton, 1960); ITT-LOCF: intent-to-treat last-observation carried-forward analysis; MIND: Minor depression; PP: per-protocol population.

Primary care providers are often confronted with such patients. If there has been pretreatment the experience obtained can guide the decision to treat or not to treat. In RCTs a high percentage of patients had earlier treatments and it is not unlikely that nonresponders are overrepresented in these study populations. Demyttenaere et al. (2008) found a significantly higher responder rate after 6 months of treatment with SSRI in patients with a first episode of major depression compared to patients with recurrent depressive episodes and previous treatment with antidepressants (88% versus 62%). Own data from the MIND-study (Hegerl et al., 2010) suggest lower sertraline–placebo differences regarding HAMD-17 sum score changes (baseline — individual endpoint) in patients with pretreatment with antidepressants (including St. John's Wort) within the last two years (ΔHAMD-17 = 1.46 points) as compared with patients who had no pretreatment with antidepressants within the last two years (ΔHAMD-17 = 2.29 points) (according to ITT-LOCF analysis). The possible benefit a patient can expect from starting for the first time an antidepressant medication would be underestimated when referring to data from typical RCTs as done by the meta-analyses by Fournier et al.(2010)and Kirsch et al.(2008). The meta-analysis by Fournier et al. (2010) illustrates how selective some meta-analyses are: The authors of this article included from the literature published in 30 years only six placebo-controlled adult outpatient antidepressant trials with only 718 patients.

3.4. High treatment responders can be picked out by a probationary treatment

A better outcome between active treatment and placebo of averaged 2 points in the HAMD-17 implies the existence of patients with an even smaller, but also those with greater benefit of the treatment. So far, markers for the identification of such "high-responders" are not available; so, it must be discussed whether or not probationary treatments are justified. Probationary treatments enable one subgroup to take advantages from a therapy which is adequate for them while treatment could be stopped in other patients. In this context, it is important to note that the decision to continue or discontinue treatment can be made already after two weeks of therapy. Several studies suggest that patients without marked symptom reduction after two weeks (e.g. <20% in HAMD-17 (Hamilton, 1960)) have a high probability to be non-responders after ten weeks of treatment (79–94%) (Henkel et al., 2009; Kemp et al., 2010; Papakostas et al., 2006; Stassen et al., 2007; Szegedi et al., 2003; Tadiç et al., 2010; Taylor et al., 2006; van Calker et al., 2009). Post-hoc analysis of data from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study points in the same direction: Early change of depressive symptoms (from baseline to week 2) was found to allow predictions of non-response after six weeks of treatment being "sufficiently certain for clinicians to spare identified patients from prolonged exposure to ineffective treatment" (Kuk et al., 2010; p. 1502). Although it will not be possible to decide how far the improvement in the individual responders is due to pharmacological effects of the antidepressant or other factors such as spontaneous remission or unspecific effects, the responder group is likely to be enriched with true and clear responders. The risks and disadvantages of such a probationary treatment are limited.

3.5. Wrong patients might be included in RCTs

Many RCTs are conducted in countries in which many patients do not own a health insurance, copayments for medical visits or prescription medication are high (e.g., USA). It becomes attractive to such uninsured and inadequately insured persons to enroll in a trial because this involves the chance to get free medical visits and medication as well as financial incentives. Moreover, inclusion of patients is mostly carried out in study centers being paid for every included patient. High time pressure is characteristic for these studies. These factors go along with a high risk that patients slip into the studies although they do not fulfill the inclusion and exclusion criteria. Especially uninsured patients might be susceptible to undue inducement in order to join a clinical trial (Pace et al., 2003) motivating them to exaggerate depression severity. After inclusion this exaggeration is not necessary any more leading to rapid initial improvements in the placebo and verum arms. It is evident that such factors will reduce the placebo–verum difference in ΔHAMD-17 (Hegerl and Mergl, 2010).

However, there are also conflictive findings: Rush et al. (2008) investigated predictors of remission to second-step medications (like sertraline and venlafaxine-XR) for treatment of major depressive disorder after intolerance or lack of remission regarding an initial selective serotonin reuptake inhibitor (citalopram hydrobromide); according to their results, patients without any insurance do not significantly differ from patients with private or public insurance in their remission rates. However, patients with only public insurance were characterized by significantly less remission rates, as compared to patients with private insurance. For the authors, this finding reflects the fact that remission is less likely in the context of social disadvantage. Another result (remission in white participants being nearly twice as likely as in nonwhite ones) points in a similar direction.

3.6. Differences in placebo effects in RCTs compared to routine care can lead to underestimation of clinical significance of treatment

Having in mind the many visits which intense studies entail, it is likely that modern RCTs induce more placebo effects due to factors such as activation, suggestion, hope induction, social support and care than what can be achieved within the small time budget of a primary care provider. Possibly, the increased efforts and complexity associated with phase-III trials might be a factor contributing to the fact that the placebo response rates in clinical antidepressant trials have increased worldwide in the last twenty years (Rief et al., 2009; Walsh et al., 2002). Other factors like increased recruitment of patients with milder forms of depression who are known to have an increased response rate to placebo have also been discussed (Silberman, 2009): These patients are given higher depression ratings at baseline than acceptable, whereas subsequent ratings are correct ("overrating") (Möller, 2008).

It may be argued that this larger placebo effect in RCTs influences outcome in the active treatment versus placebo group in a similar manner and may therefore have no net effect on observed differences in ΔHAMD-17 or other outcome parameters. This is, however, only partly true, because in the active treatment arm nonspecific and true pharmacological effects of the antidepressant are not additive. Some additional responders to nonspecific effects would also have responded to the antidepressant alone.

In this context, a further factor comes into play. Because of blinding, in RCT efficacy studies hope is induced in a similar manner in both the active treatment and placebo arms. This is not the situation found in routine care. If the doctor decides not to offer specific treatment, but proposes watchful waiting, less placebo (placebo = "I will please") effects can be expected. Some patients expecting help might even respond with deception and nocebo effects cannot be excluded. In addition, if the doctor decides to start antidepressant treatment more hope might be induced than in the RCT. The patient in daily praxis knows that he gets a drug which is supposed to work, whereas in RCTs the patient knows he has a considerable risk to receive only a placebo. Less hope is induced by the active treatment in RCTs and the improvement might be smaller than that in daily practice. The size of this reduced hope induction in RCTs seems to be important as suggested by a recent metaanalytic study (Sneed et al., 2008): Antidepressant response rates were 60% in placebo-controlled RCTs with three arms and two active drugs (chance to get verum: 66%), but only 46% in placebo-controlled two-arm RCTs (chance to get verum: 50%). Analogously, a more recent study (Sinyor et al., 2010) suggests a strong influence of the number of active treatment arms in a clinical trial on the placebo response rates. Moreover, time is an important variable in the evaluation of antidepressant versus placebo response (e.g., Frank et al., 1990; Montgomery et al., 1993).

Altogether, these factors reduce the placebo–verum difference in ΔHAMD-17 compared to what a patient can expect when receiving an antidepressant instead of "watchful waiting" (see Fig. 2).

3.7. In RCTs the possibilities to adjust treatment individually are limited

In daily practice, guidelines recommend changing treatment with antidepressants after two to four weeks if no clinical improvement is observed. Adjustment of dosage, changing to another antidepressant with different pharmacodynamic properties or augmentation strategies are good clinical practice. Although not proven by stringent studies, it is likely that these established strategies increase the probability that the right dosage and the right drug will be found for the individual patient, and the chance to respond will be increased as compared to RCT in which one antidepressant is used for up to ten weeks even if no improvement is seen. In daily practice also patient preference can be better taken into account. Within the MIND-study patients receiving antidepressant drug treatment in line with their treatment preference had a better outcome by statistical trend than those who received an antidepressant, but preferred psychotherapy (outcome difference 2.86 points in HAMD-17; $p = 0.07$) (Mergl et al., 2011).

3.8. Antidepressants have the additional benefit of prevention of relapse

Antidepressant drugs have the additional benefit for the patient that they have clearly proven and strong effects concerning prevention of relapse and recurrence (e.g., Blier et al., 2007; Geddes et al., 2006; Kornstein, 2008; Reynolds

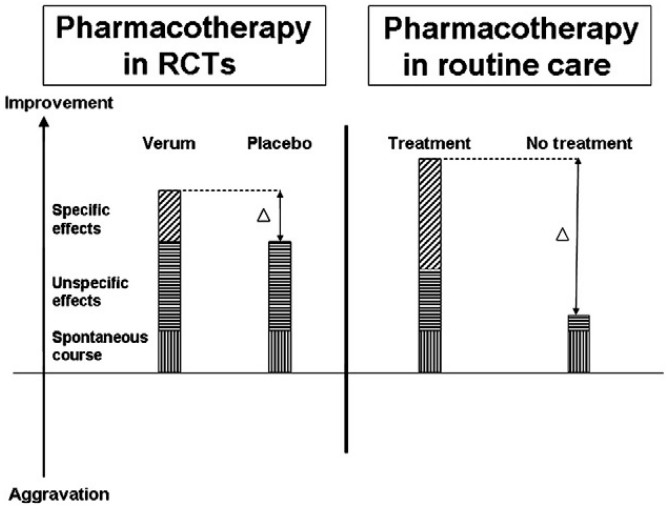

Fig. 2. A simple model for spontaneous course, unspecific effects (e.g. induction of hope, activation and care) and specific effects of antidepressants as additive factors contributing to treatment response. The response difference (ΔHAMD-17) between antidepressants and placebo within RCTs can be expected to be clearly smaller than that between antidepressants and watchful waiting in daily practice. In the latter case, the effect of antidepressants is likely to be even larger than in RCTs and the effect of watchful waiting to be less pronounced than that of placebo.Notes: Δ: Placebo–verum response difference; HAMD-17: Hamilton Depression Rating Scale (17-item version) (Hamilton, 1960) sum score; RCT: Randomized controlled trial.

et al., 2006). Compared with treatment discontinuation, antidepressants can reduce the risk of relapse or recurrence by about 50–70%. The mean rate of relapse on placebo was 41%, as compared to 18% on antidepressants (Geddes et al., 2003). Focusing only on the acute effects neglects this additional advantage of an antidepressant for the patients.

3.9. To define an improvement in HAMD-17 of two points is questionable

Independent of the arguments presented so far, defining two points in the HAMD-17 as not being clinically relevant is questionable. Such an improvement can, for example simply signify the change from "wishes s/he were dead" to "suicidality absent" (Hamilton, 1960) or a clear improvement in appetite or sleep.

4. Clinical significance of effects of psychotherapy as a possible alternative to antidepressants

Regarding treatment of mild depression, some guidelines prefer psychotherapy (e.g., the German guidelines for the treatment of depressive disorders (DGPPN et al., 2009)). Also in the recent reviews about clinical significance of antidepressants in mild depression, skeptical opinions about antidepressants are often combined with support for psychotherapy as a better alternative (e.g., Kirsch et al., 2008). This happens in spite of the fact that for psychotherapy the evidence base for efficacy is considerably more difficult to demonstrate than that for antidepressants. This results mainly from difficulties in defining a control group and in blinding therapists as well as patients (e.g., Nutt and Sharpe, 2008). Without blinding of patients, no placebo effect will be induced in the control group. Knowing that one is only in a 'waiting list group', 'treatment as usual group' or a 'moderated self-help group' might be frustrating and could even result in nocebo effects. This was shown by the above mentioned MIND-study (Hegerl et al., 2010). The striking finding was that the outcome in the moderated self-help group as a psychotherapy control condition was not only significantly worse compared with the Cognitive Behavioral Therapy group, but also compared with the pill placebo group and all other groups (see Fig. 3). Demonstrating superiority over such a psychotherapy control group is obviously not providing the same level of evidence for efficacy as pill–placebo–antidepressant comparisons (Hegerl et al., 2010). If patients know that

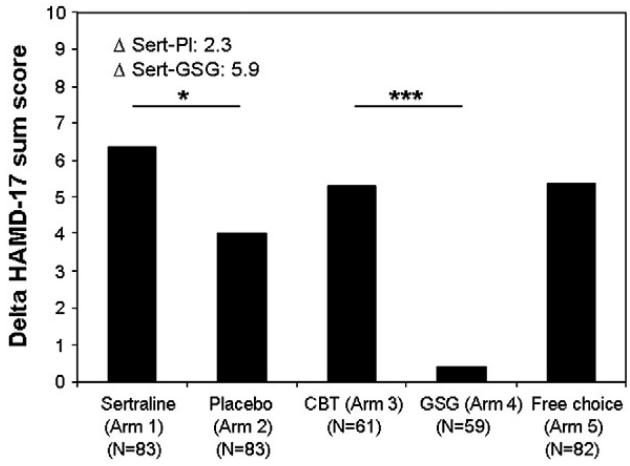

Fig. 3. HAMD-17 sum score changes (baseline — endpoint) in patients with mild-to-moderate depressive disorders (according to MMRM analysis). Analysis was based on data from the MIND study (Hegerl et al., 2010). The guided self-help groups were found to lead to a significantly worse outcome than the other four arms of the MIND study including pill–placebo.Notes: CBT: Cognitive–behavioral group therapy; free choice: free choice of either sertraline or CBT; GSG: Guided self-help groups; HAMD-17: Hamilton Depression Rating Scale — 17-item version (Hamilton, 1960); MIND: Minor depression; MMRM: Mixed-model repeated-measures analysis (Twisk, 2003); N: sample size; Pl: Placebo; Sert: Sertraline.

they had the bad luck to be only in the control condition the consequence is likely to be frustration, despair and nonadherence (nocebo instead of placebo effects). The MIND study suggests such effects despite the control condition having involved unspecific support, psychoeducational and stress management elements. The scope of this methodological problem is not always sufficiently recognized and considered in psychotherapy research. It explains why the effect sizes of psychotherapies are especially large in psychotherapy studies involving waiting list control groups which are likely to have especially high nocebo effects (Robinson et al., 1990). It is bad news for every depressed patient seeking help to be randomized to a waiting list control group. Such groups are not a valid control condition and should not be included in meta analyses.

In addition to this fundamental problem in psychotherapy research, a recent analysis indicates that many studies have further methodological shortcomings (no clear diagnostic criteria for depressive disorders, no treatment manual, no training of psychotherapists, no checking of treatment integrity, no ITT analyses, small sample sizes (<50), no randomization by an independent party, no blinding of outcome assessors) leading to an overestimation of treatment effects of psychotherapy (Cuijpers et al., 2010).

5. What to do?

If the drug–placebo difference in depression ratings cannot be used to judge the clinical relevance of antidepressant effects in daily practice, what might be a better alternative? The risk to discard helpful treatment has to be minimized. Thus, for assessing clinical significance it is recommended

- to use per protocol analyses;
- to install strict compliance control;
- to include mainly patients without pre-treatment;
- to consider other factors such as patients' preferences (Mergl et al., 2011), age, comorbidity, comedication, earlier treatment experiences as inclusion/exclusion criteria;
- to use open designs (but blinded raters) with randomization at the center/practice level and
- to compare the improvement with antidepressants not to pill–placebo, but to active monitoring.

6. Conclusion

The paper focused on factors possibly leading to an underestimation of the clinical significance of antidepressant treatment in daily practice when measured with the presently used approach (placebo–verum difference in HAMD-17). It did not address factors which might introduce a bias in the opposite direction, leading to an overestimation of the clinical significance of treatment effects in daily practice. Among these are problems with blinding in RCTs (Porter et al., 2003; Ventegodt et al., 2009) and exclusion of patients with comorbidity and comedication (Mosenifar, 2007; Van Spall et al., 2007) which might suggest better efficacy than that observable in daily practice. Such factors, however, do not invalidate the presented arguments which highlight the risk to reject clearly effective antidepressant treatments for patients with mild depression when relying on the presently used assessment of clinical significance.

Role of funding source

There was no study sponsor involved in writing of this review article.

Conflict of interest

The authors have no relevant financial relationships to disclose.

Acknowledgments

We would like to thank all investigators and patients who participated in the MIND study.

References

Akerblad, A.C., Bengtsson, F., Ekselius, L., von Knorring, L., 2003. Effects of an educational compliance enhancement programme and therapeutic drug monitoring on treatment adherence in depressed patients managed by general practitioners. Int. Clin. Psychopharmacol. 18, 347–354.

Angst, J., Merikangas, K., 1997. The depressive spectrum: diagnostic classification and course. J. Affect. Disord. 45, 31–39 discussion 39–40.

Angst, J., Delini-Stula, A., Stabl, M., Stassen, H.H., 1993. Is a cut-off score a suitable measure of treatment outcome in short-term trials in depression? A methodological meta-analysis. Hum. Psychopharmacol. 8, 311–317.

APA, 2000a. Practice guideline for the treatment of patients with major depressive disorder (revision). American Psychiatric Association. Am. J. Psychiatry 157, 1–45.

APA, 2000b. Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR (Text Revision). American Psychiatric Publication Inc., Washington DC.

Bauer, M., Bschor, T., Pfennig, A., Whybrow, P.C., Angst, J., Versiani, M., Möller, H.-J., 2007. World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for biological treatment of unipolar depressive disorders in primary care. World J. Biol. Psychiatry 8, 67–104.

Bech, P., 2006. Rating scales in depression: limitations and pitfalls. Dialogues Clin. Neurosci. 8, 207–215.

Beck, A.T., 1967. Depression: Clinical, Experimental, and Theoretical Aspects. University of Pennsylvania Press, Philadelphia, Pennsylvania.

Blier, P., Keller, M.B., Pollack, M.H., Thase, M.E., Zajecka, J.M., Dunner, D.L., 2007. Preventing recurrent depression: long-term treatment for major depressive disorder. J. Clin. Psychiatry 68, e06.

Cuijpers, P., Smit, F., 2002. Excess mortality in depression: a meta-analysis of community studies. J. Affect. Disord. 72, 227–236.

Cuijpers, P., Smit, F., Oostenbrink, J., de Graaf, R., Ten Have, M., Beekman, A., 2007. Economic costs of minor depression: a population-based study. Acta Psychiatr. Scand. 115, 229–236.

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S.D., Andersson, G., 2010. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. Psychol. Med. 40, 211–223.

Davidson, J.R., 2010. Major depressive disorder treatment guidelines in America and Europe. J. Clin. Psychiatry 71 (Suppl. E1), e04.

de Maat, S.M., Dekker, J., Schoevers, R.A., de Jonghe, F., 2007. Relative efficacy of psychotherapy and combined therapy in the treatment of depression: a meta-analysis. Eur. Psychiatry 22, 1–8.

Demyttenaere, K., Adelin, A., Patrick, M., Walthere, D., Katrien de, B., Michele, S., 2008. Six-month compliance with antidepressant medication in the treatment of major depressive disorder. Int. Clin. Psychopharmacol. 23, 36–42.

DGPPN, BÄK, KBV, AWMF, AkdÄ, BPtK, BApK, DAGSHG, DEGAM, DGPM, DGPs, DGRW, 2009. S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression - Langfassung. DGPPN, ÄZQ, AWMF, Berlin, Düsseldorf.

Faries, D., Herrera, J., Rayamajhi, J., DeBrota, D., Demitrack, M., Potter, W.Z., 2000. The responsiveness of the Hamilton Depression Rating Scale. J. Psychiatr. Res. 34, 3–10.

Fergusson, D.M., Horwood, L.J., Ridder, E.M., Beautrais, A.L., 2005. Subthreshold depression in adolescence and mental health outcomes in adulthood. Arch. Gen. Psychiatry 62, 66–72.

Fogel, J., Eaton, W.W., Ford, D.E., 2006. Minor depression as a predictor of the first onset of major depressive disorder over a 15-year follow-up. Acta Psychiatr. Scand. 113, 36–43.

Fountoulakis, K.N., Möller, H.-J., 2010. Efficacy of antidepressants: a re-analysis and re-interpretation of the Kirsch data. Int. J. Neuropsychopharmacol. 1–8.

Fournier, J.C., DeRubeis, R.J., Hollon, S.D., Dimidjian, S., Amsterdam, J.D., Shelton, R.C., Fawcett, J., 2010. Antidepressant drug effects and depression severity: a patient-level meta-analysis. JAMA 303, 47–53.

Frank, E., Kupfer, D.J., Perel, J.M., Cornes, C., Jarrett, D.B., Mallinger, A.G., Thase, M.E., McEachran, A.B., Grochocinski, V.J., 1990. Three-year outcomes for maintenance therapies in recurrent depression. Arch. Gen. Psychiatry 47, 1093–1099.

Geddes, J.R., Carney, S.M., Davies, C., Furukawa, T.A., Kupfer, D.J., Frank, E., Goodwin, G.M., 2003. Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. Lancet 361, 653–661.

Geddes, J., Freemantle, N., Mason, J., Eccles, M., Boynton, J., 2006. Selective serotonin reuptake inhibitors (SSRIs) versus other antidepressants for depression. Cochrane Database Syst. Rev. 3, CD001851.

Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23, 56–62.

Hegerl, U., Mergl, R., 2010. The clinical significance of antidepressant treatment effects cannot be derived from placebo–verum response differences. J. Psychopharmacol. 24, 445–448.

Hegerl, U., Hautzinger, M., Mergl, R., Kohnen, R., Schütze, M., Scheunemann, W., Allgaier, A.-K., Coyne, J., Henkel, V., 2010. Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: a randomized, controlled trial including a patients' choice arm. Int. J. Neuropsychopharmacol. 13, 31–44.

Henkel, V., Seemüller, F., Obermeier, M., Adli, M., Bauer, M., Mundt, C., Brieger, P., Laux, G., Bender, W., Heuser, I., Zeiler, J., Gaebel, W., Mayr, A., Möller, H.-J., Riedel, M., 2009. Does early improvement triggered by antidepressants predict response/remission? Analysis of data from a naturalistic study on a large sample of inpatients with major depression. J. Affect. Disord. 115, 439–449.

International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2001. ICH Harmonised Tripartite Guideline for Good Clinical Practice: With EC Directive (ICH step 4 guidelines). Brookwood Medical Publications, Richmond on Thames.

Judd, L.L., Akiskal, H.S., Maser, J.D., Zeller, P.J., Endicott, J., Coryell, W., Paulus, M.P., Kunovac, J.L., Leon, A.C., Mueller, T.I., Rice, J.A., Keller, M.B., 1998. A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders. Arch. Gen. Psychiatry 55, 694–700.

Kemp, D.E., Calabrese, J.R., Eudicone, J.M., Ganocy, S., Tran, Q.V., McQuade, R.D., Marcus, R.N., Vester-Blokland, E., Owen, R., Carlson, B.X., 2010. Predictive value of early improvement in bipolar depression trials: a post-hoc pooled analysis of two 8-week aripiprazole studies. Psychopharmacol. Bull. 43, 5–27.

Kirsch, I., Deacon, B.J., Huedo-Medina, T.B., Scoboria, A., Moore, T.J., Johnson, B.T., 2008. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med. 5, e45.

Klinkman, M.S., Coyne, J.C., Gallo, S., Schwenk, T.L., 1997. Can case-finding instruments be used to improve physician detection of depression in primary care? Arch. Fam. Med. 6, 567–573.

Kornstein, S.G., 2008. Maintenance therapy to prevent recurrence of depression: summary and implications of the PREVENT study. Expert. Rev. Neurother. 8, 737–742.

Kuk, A.Y., Li, J., Rush, A.J., 2010. Recursive subsetting to identify patients in the STAR*D: a method to enhance the accuracy of early prediction of treatment outcome and to inform personalized care. J. Clin. Psychiatry 71, 1502–1508.

Lyness, J.M., Heo, M., Datto, C.J., Ten Have, T.R., Katz, I.R., Drayer, R., Reynolds III, C.F., Alexopoulos, G.S., Bruce, M.L., 2006. Outcomes of minor and subsyndromal depression among elderly patients in primary care settings. Ann. Intern. Med. 144, 496–504.

Mergl, R., Henkel, V., Allgaier, A.-K., Kramer, D., Hautzinger, M., Kohnen, R., Coyne, J.C., Hegerl, U., 2011. Are treatment preferences relevant for response to serotonergic antidepressants and cognitive–behavioural therapy in depressed primary care patients? Results from a randomized controlled trial including a patients' choice arm. Psychother. Psychosom. 80, 39–47.

Middleton, H., Shaw, I., Hull, S., Feder, G., 2005. NICE guidelines for the management of depression. BMJ 330, 267–268.

Möller, H.-J., 2008. Isn't the efficacy of antidepressants clinically relevant? A critical comment on the results of the metaanalysis by Kirsch et al. 2008. Eur. Arch. Psychiatry Clin. Neurosci. 258, 451–455.

Montgomery, S.A., Rasmussen, J.G., Tanghøj, P., 1993. A 24-week study of 20 mg citalopram, 40 mg citalopram, and placebo in the prevention of relapse of major depression. Int. Clin. Psychopharmacol. 8, 181–188.

Mosenifar, Z., 2007. Population issues in clinical trials. Proc. Am. Thorac. Soc. 4, 185–187 discussion 187–188.

NICE, 2004. Depression: management of depression in primary and secondary care. Clinical Practice Guideline No 23. National Institute for Clinical Excellence, London.

NICE, 2009. Depression: the treatment and management of depression in adults. National Clinical Practice Guideline 90. National Institute for Health and Clinical Excellence, London.

Nierenberg, A.A., Rapaport, M.H., Schettler, P.J., Howland, R.H., Smith, J.A., Edwards, D., Schneider, T., Mischoulon, D., 2010. Deficits in psychological well-being and quality-of-life in minor depression: implications for DSM-V. CNS Neurosci. Ther. 16, 208–216.

Nutt, D.J., Sharpe, M., 2008. Uncritical positive regard? Issues in the efficacy and safety of psychotherapy. J. Psychopharmacol. 22, 3–6.

Pace, C., Miller, F.G., Danis, M., 2003. Enrolling the uninsured in clinical trials: an ethical perspective. Crit. Care Med. 31, S121–S125.

Pampallona, S., Bollini, P., Tibaldi, G., Kupelnick, B., Munizza, C., 2002. Patient adherence in the treatment of depression. Br. J. Psychiatry 180, 104–109.

Papakostas, G.I., Perlis, R.H., Scalia, M.J., Petersen, T.J., Fava, M., 2006. A meta-analysis of early sustained response rates between antidepressants and placebo for the treatment of major depressive disorder. J. Clin. Psychopharmacol. 26, 56–60.

Porter, R., Frampton, C., Joyce, P.R., Mulder, R.T., 2003. Randomized controlled trials in psychiatry. Part 1: methodology and critical evaluation. Aust. N. Z. J. Psychiatry 37, 257–264.

Rapaport, M.H., Judd, L.L., 1998. Minor depressive disorder and subsyndromal depressive symptoms: functional impairment and response to treatment. J. Affect. Disord. 48, 227–232.

Reynolds III, C.F., Dew, M.A., Pollock, B.G., Mulsant, B.H., Frank, E., Miller, M.D., Houck, P.R., Mazumdar, S., Butters, M.A., Stack, J.A., Schlernitzauer, M.A., Whyte, E.M., Gildengers, A., Karp, J., Lenze, E., Szanto, K., Bensasi, S., Kupfer, D.J., 2006. Maintenance treatment of major depression in old age. N. Engl. J. Med. 354, 1130–1138.

Rief, W., Nestoriuc, Y., Weiss, S., Welzel, E., Barsky, A.J., Hofmann, S.G., 2009. Meta-analysis of the placebo response in antidepressant trials. J. Affect. Disord. 118, 1–8.

Robinson, L.A., Berman, J.S., Neimeyer, R.A., 1990. Psychotherapy for the treatment of depression: a comprehensive review of controlled outcome research. Psychol. Bull. 108, 30–49.

Rush, A.J., Wisniewski, S.R., Warden, D., Luther, J.F., Davis, L.L., Fava, M., Nierenberg, A.A., Trivedi, M.H., 2008. Selecting among second-step antidepressant medication monotherapies: predictive value of clinical, demographic, or first-step treatment features. Arch. Gen. Psychiatry 65, 870–880.

Shelton, R.C., Prakash, A., Mallinckrodt, C.H., Wohlreich, M.M., Raskin, J., Robinson, M.J., Detke, M.J., 2007. Patterns of depressive symptom response in duloxetine-treated outpatients with mild, moderate or more severe depression. Int. J. Clin. Pract. 61, 1337–1348.

Silberman, S., 2009. Placebos are getting more effective. Drugmakers are desperate to know why. Wired Magazine (08.24.09). http://www.wired.com/medtech/drugs/magazine/17-09/ff_placebo_effect?currentPage=all 2009.

Sinyor, M., Levitt, A.J., Cheung, A.H., Schaffer, A., Kiss, A., Dowlati, Y., Lanctot, K.L., 2010. Does inclusion of a placebo arm influence response to active antidepressant treatment in randomized controlled trials? Results from pooled and meta-analyses. J. Clin. Psychiatry 71, 270–279.

Sneed, J.R., Rutherford, B.R., Rindskopf, D., Lane, D.T., Sackeim, H.A., Roose, S.P., 2008. Design makes a difference: a meta-analysis of antidepressant response rates in placebo-controlled versus comparator trials in late-life depression. Am. J. Geriatr. Psychiatry 16, 65–73.

Stassen, H.H., Angst, J., Hell, D., Scharfetter, C., Szegedi, A., 2007. Is there a common resilience mechanism underlying antidepressant drug response? Evidence from 2848 patients. J. Clin. Psychiatry 68, 1195–1205.

Stolk, P., Ten Berg, M.J., Hemels, M.E., Einarson, T.R., 2003. Meta-analysis of placebo rates in major depressive disorder trials. Ann. Pharmacother. 37, 1891–1899.

Szegedi, A., Müller, M.J., Anghelescu, I., Klawe, C., Kohnen, R., Benkert, O., 2003. Early improvement under mirtazapine and paroxetine predicts later stable response and remission with high sensitivity in patients with major depression. J. Clin. Psychiatry 64, 413–420.

Tadiç, A., Helmreich, I., Mergl, R., Hautzinger, M., Kohnen, R., Henkel, V., Hegerl, U., 2010. Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. J. Affect. Disord. 120, 86–93.

Task Force for the APA Handbook of Psychiatric Measures, 2000. Handbook of Psychiatric Measures. APA, Washington DC.

Taylor, M.J., Freemantle, N., Geddes, J.R., Bhagwagar, Z., 2006. Early onset of selective serotonin reuptake inhibitor antidepressant action: systematic review and meta-analysis. Arch. Gen. Psychiatry 63, 1217–1223.

Twisk, J.W.R., 2003. Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide. Cambridge University Press, New York.

van Calker, D., Zobel, I., Dykierek, P., Deimel, C.M., Kech, S., Lieb, K., Berger, M., Schramm, E., 2009. Time course of response to antidepressants: predictive value of early improvement and effect of additional psychotherapy. J. Affect. Disord. 114, 243–253.

Van Spall, H.G., Toren, A., Kiss, A., Fowler, R.A., 2007. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. JAMA 297, 1233–1240.

Ventegodt, S., Andersen, N.J., Brom, B., Merrick, J., Greydanus, D.E., 2009. Evidence-based medicine: four fundamental problems with the randomized clinical trial (RCT) used to document chemical medicine. Int. J. Adolesc. Med. Health 21, 485–496.

Vuorilehto, M., Melartin, T., Isometsa, E., 2005. Depressive disorders in primary care: recurrent, chronic, and co-morbid. Psychol. Med. 35, 673–682.

Walsh, B.T., Seidman, S.N., Sysko, R., Gould, M., 2002. Placebo response in studies of major depression: variable, substantial, and growing. JAMA 287, 1840–1847.

WHO, 1992. The ICD-10 Classification of Mental and Behavioural Disorders. Clinical Descriptions and Diagnostic Guidelines. World Health Organization, Geneva.