Journal of Psychiatric Research 44 (2010) 1063–1068

Contents lists available at ScienceDirect

Journal of Psychiatric Research

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

Response and remission criteria in major depression – A validation of current practice☆

Michael Riedel [a], Hans-Jürgen Möller [a], Michael Obermeier [a], Rebecca Schennach-Wolff [a], Michael Bauer [b], Mazda Adli [c], Klaus Kronmüller [d], Thomas Nickel [e], Peter Brieger [f], Gerd Laux [g], Wolfram Bender [h], Isabella Heuser [i], Joachim Zeiler [j], Wolfgang Gaebel [k], Florian Seemüller [a,*]

[a] Department of Psychiatry and Psychotherapy, Ludwig-Maximilians-University Munich, Nussbaumstrasse 7, 80336 Munich, Germany
[b] Department of Psychiatry and Psychotherapy, Carl Gustav Carus University Dresden, Technical University Dresden, Fetscherstr. 74, 01307 Dresden, Germany
[c] Department of Psychiatry and Psychotherapy, Campus, Charité Mitte (CCM), Charitéplatz 1, 10117 Berlin, Germany
[d] Department of Psychiatry and Psychotherapy, University of Heidelberg, Voßstr. 2, 69115 Heidelberg, Germany
[e] Max Planck Institute of Psychiatry, Kraepelinstr. 2-7, 80804 Munich, Germany
[f] Department of Psychiatry and Psychotherapy, Martin-Luther University Halle-Wittenberg, Julius-Kühn-Str.7, 06097 Halle, Germany
[g] Department of Psychiatry and Psychotherapy, Inn-Salzach-Klinikum. Garbersee 7, 83512 Wasserburg, Germany
[h] Department of Psychiatry and Psychotherapy, Isar-Amper-Klinikum Munich East, Vockestr. 72, 85540 Haar, Germany
[i] Department of Psychiatry and Psychotherapy, Campus Charité Benjamin Franklin (CFB), Eschenallee 3, 14050 Berlin, Germany
[j] Department of Psychiatry and Psychotherapy, Auguste-Viktoria-Krankenhaus, Rubensstr. 125, 12157 Berlin, Germany
[k] Department of Psychiatry and Psychotherapy, University of Düsseldorf, Bergische Landstr. 2, 40629 Düsseldorf, Germany

ARTICLE INFO

Article history:
Received 1 December 2009
Received in revised form
4 March 2010
Accepted 16 March 2010

Keywords:
Remission
Response
Major depression
Bipolar
Cut-off
Inpatients
HAMD-17
HMD-21
MADRS
CGI
CGI
BDI
BDI

ABSTRACT

Remission and response were suggested as the most relevant outcome criteria for the treatment of depression. There is still marked uncertainty as to what cut-offs should be used on current depression rating scales. The goal of the present study was to compare the validity of different HAMD, MADRS and BDI cut-offs for response and remission.

The naturalistic prospective study was performed in 12 psychiatric hospitals in Germany. All evaluable patients ($n = 846$) were hospitalized and had to meet DSM-IV criteria for major depressive disorder. Biweekly ratings were assessed using HAMD-21, MADRS and BDI. A CGI-S score of 1 and a CGI-I score of at least 2 was used as the primary comparative measure of remission and response, respectively.

A HAMD-21 cut-off ≤ 7 (AUC: 0.92), HAMD-17 cut-of ≤ 6 (AUC: 0.90), MADRS cut-off ≤ 7 (AUC: 0.94) and BDI cut-off ≤ 12 (AUC: 0.83) were associated with a maximum of specificity and sensitivity for defining remission.

A minimum decrease of 47% of the HAMD-21 (AUC: 0.90), $\leq$57% for HAMD-17 (AUC: 0.89), $\leq$ 46% for MADRS (0.91) and a decrease of 47% for the BDI baseline score (AUC: 0.78) best corresponded CGI response criteria.

Our data largely confirmed currently used remission and response criteria in naturalistically treated patients.

1. Introduction

Leading international drug authorities like the FDA and the EMEA require significant drug placebo differences in the primary endpoint of at least two randomized placebo-controlled trials before allowing a new drug to be sold on the market. Also, the use of standardized rating scales is strongly recommended and regulated by the FDA within the Good Clinical Practise Guidelines (GCP). The most widely used and thus the gold standard for the assessment of depressive symptoms is the Hamilton depression rating

☆ The study was performed within the framework of the German Research Network on Depression, which was funded by the German Federal Ministry for Education and Research BMBF (01GI0219). The BMBF had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

* Corresponding author. Tel.: +49 89 5160 5846; fax: +49 89 5160 5774.
E-mail address: florian.seemueller@med.uni-muenchen.de (F. Seemüller).

0022-3956/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jpsychires.2010.03.006

scale (Hamilton, 1967). It was primarily developed for inpatients with major depression, who tend to present with melancholic features, but has been extensively used in outpatient studies as well. Unfortunately, in the meantime there are many different versions (HAMD-24, -27 and -29) of the HAMD, the two most widely used being the HAMD-21-item and the 17-item versions as originally recommended by Hamilton (Hamilton, 1967).

Its main rival is the MADRS (Montgomery and Asberg, 1979) which may possess an even better sensitivity for detecting symptom change. With respect to self- ratings, the BDI has achieved wide acceptance (Schwab et al., 1967).

The most commonly used analytical method is to compare baseline and endpoint mean scores, a procedure which lacks the information of clinical significance and the information of the illness course. One established approach is to define categories of clinical significance. In depression research the most widely used are response and remission. Unfortunately there are many different, divergent and contradictory definitions still in circulation, which often do not differentiate between different rating scale versions, thus hindering comparisons across studies. For the HAMD, for example, the definition of remission varies between thresholds <6 and ≤ 10 for the HAMD-21 and HAMD-17-item scale (Nierenberg and DeCecco, 2001). For response, definitions usually include a percental difference from the respective initial baseline score starting from $\geq 25\%$ and going up to $\geq 50\%$, but sometimes they also include a numerical threshold such as a HAMD < 10 (Riso et al., 1997). So the bizarre situation emerges that one investigator's remission is another investigator's response. Compounding the problem, major depression itself is a very inhomogeneous illness with many subtypes and a highly heterogeneous illness course. So in the end the results of different depression trials are hardly comparable. Establishing and evaluating distinct study endpoints would eliminate one factor contributing to the enormous outcome variance of antidepressant treatment trials. The task force of the Mack Arthur Foundation Research Network on the Psychobiology of Depression tried to find consistent criteria for remission, response, recovery and relapse in major depression (Frank et al., 1991; Prien et al., 1991). Even for the most frequently used terms "response, remission, recovery and relapse" no acceptable operationally defined criteria that could be used in research were found (Prien et al., 1991).

The CGI (Clinical Global Impression) has two basic scales covering disease severity (the CGI-Severity or CGI-S) and treatment-induced disease improvement (the CGI-Improvement or CGI-I). In contrast to psychopathological rating scales, it consists solely of one single item covering overall illness severity and improvement on a seven-point Likert scale. The CGI-S requires the clinician to rate the overall severity of the patient's illness at the assessment time in relation to the clinician's past experience with patients having the same diagnosis on a scale between 1 and 7 (ranging from 1 = "normal, not ill" to 7 = "extremely ill"). The CGI-I captures the overall improvement relative to baseline ranging from (1 = "very much improved" to 7 = "very much worse"). This scale might thus capture illness severity and therapeutic improvement from a different one-dimensional global perspective, compared to differentiated psychopathological rating scales. Furthermore, the CGI has been shown to be a reliable and valid measure of disease severity and to be sensitive to change (Guy, 1976). The CGI has the main advantage that outcome constructs like remission and response are very easily transferable to e.g. a CGI-S score of 1 and CGI-I score ≤ 2 (at least "much improved"), whereas HAMD, MADRS and BDI do not have predefined thresholds for response or remission (Bandelow et al., 2006).

In line with a recent suggestion made by Berk and colleagues we therefore chose to use the CGI as a validation criterion, analyzing data from a large naturalistic trial on 843 inpatients with major depressive episode who were assessed every second week until discharge. For evaluation of valid cut-offs for response and remission Berk and colleagues associated the corresponding mean values of the Young Mania Rating Scale (YMRS) and the MADRS in bipolar patients with a CGI-Severity of 1 for remission and a CGI-Improvement of at least two for response. We aimed to empirically verify remission and response criteria within a sample of depressed inpatients for the HAMD-21, HAMD-17, MADRS and BDI against the CGI using the same thresholds (Berk et al., 2008) computing receiver operating curves analysis (ROC) and applying bootstrap techniques.

2. Method

2.1. Sample and data collection

The main objective and details of the study protocol are described in detail elsewhere (Seemuller et al., 2010). In brief, data from a large prospective, naturalistic, multicenter study ($N = 1014$) were analyzed. The study was part of the German research network, funded by the German Federal Ministry of Education and Research (BMBF). Subjects were recruited from seven German psychiatric university or research hospitals (two in Munich, two in Berlin, Tübingen, Düsseldorf, Halle) and five psychiatric district hospitals (Munich, Gabersee, and three in Berlin).

Inclusion criteria were age between 18 and 65 and signed written informed consent. Patients had to meet ICD-10 diagnostic criteria for any major depressive episode (ICD-10: F31.3x–5x, F32, F33, F34, F38, F39) or for a depressive disorder not otherwise specified (ICD-10). Moreover, for confirmation of the diagnose of a depressive spectrum disorder according to DSM-IV as well as for the detection of relevant axis I and axis II comorbidities, the Structured Clinical Interview for DSM-IV (SCID-I and SCID-II) was used (Wittchen et al., 1997).

Psychopathological symptoms were assessed using the Hamilton Depression Rating Scale (HAMD-21) (Baumann, 1976). Its German 17-item version has shown good reliability with a Cornbachs α ranging from 0.72–0.83 (Baumann, 1976; Maier et al., 1985).

The German translation of the Montgomery Asberg Depression Rating Scale (MADRS) (Schmidtke et al., 1988) has been shown to have a high internal consistency (Cornbachs $\alpha = 0.86$) and a high sensitivity for change (Schmidtke et al., 1988). Its validity has been demonstrated by moderate to good correlations with the 17-item German version of the HAMD ranging from 0.51 to 0.89 (Schmidtke et al., 1988).

The German version of the self-rated Becks Depression Inventory (BDI) (Hautzinger, 1991) has a similar internal consistency (Cornbachs $\alpha = 0.86$), good correlations with the self-rated Zungs Depression scale and moderate to poor correlations with the HAMD (Pearson correlation = 0.37) (Hautzinger, 1991).

Ratings were assessed by clinicians who had undergone a minimum of four years' clinical training in psychiatry. All ratings for each patient were assessed by the same clinician. Patients were rated according to the protocol at baseline and every two weeks until discharge. Patients were included in the analysis if at least two assessments were available.

2.2. Treatment

Patients were treated at the discretion of the psychiatrist in charge under consideration of the international clinical guidelines for the treatment of depression (APA, WSFBP) (Bauer et al., 2007; American Psychiatric Association, 2000; Deutsche Gesellschaft für