Chapter 3

The generalizability of antidepressant efficacy trials to routine psychiatric outpatient practice

Rosalind van der Lem
Nic J.A. van der Wee
Tineke van Veen
Frans G. Zitman

Psychological Medicine 2011:41:1353-1363

ABSTRACT

Background: Generalizability of antidepressants efficacy trials (AETs) to daily practice is questioned because of their very stringent patient selection. This study aims to determine eligibility for AETs of outpatients suffering from major depression in a routine outpatient-setting and investigates influence of eligibility on treatment outcome.

Methods: Data collection (n=1653) through routine outcome monitoring by independent trained research nurses. MINIplus and DAPP-SF were used for diagnostic assessment and personality pathology screening. MADRS was used for assessment of baseline severity and treatment outcome. Eligibility was assessed by stepwise application of commonly used exclusion criteria. Influence of eligibility on treatment outcome was investigated in a subsample of the 1653 patients who had at least one follow up assessment (n=626). Eligible and non-eligible patients were compared on proportion of response (50% reduction) and remission on MADRS (MADRS≤10).

Results: 17–25% of the patients were eligible for AETs. The most common reasons for exclusion would be "not meeting minimum baseline severity" and "presence of co morbid Axis I disorder". Eligible and non-eligible patients did not differ in treatment outcome. Only "meeting the minimum baseline severity" is associated with remission.

Conclusion: The majority of "real life" outpatients is not eligible for AETs. However, the influence of eligibility on treatment outcome seems to be small. This suggests that stringent patient selection by eligibility criteria is not the major reason for lack of generalizability of AETs. Exclusion of less severely depressed patients from the analyses resulted in better treatment outcome. Milder depression is highly prevalent in daily practice and more research into treatment effectiveness in milder depression is warranted.

Key words: major depression; routine outcome monitoring; generalizability; antidepressant efficacy trials; eligibility

INTRODUCTION

During the past decades, depression treatment has shifted from an approach based on clinical expertise towards an evidence based approach using results from randomized clinical trials (RCTs) on antidepressants and/or psychotherapy [1,2]. However, for methodological and ethical reasons, antidepressant efficacy trials (AETs) will always need strict, randomized and placebo-controlled conditions, and use stringent inclusion and exclusion criteria for patient selection. In this way, internal validity is optimized. However, by optimizing internal validity, external validity (generalizability) might be compromised. Hence, the generalizability of the results from AETs to clinical practice can be questioned [3-7]. Three studies in the United States examined the eligibility of depressive patients for inclusion in AETs [8-10]. These reported that only 12–34% of these patients were eligible for AETs. However, these investigations regarded only fee-for-service settings, which may not be generalizable to the European healthcare system. In a European study, the eligibility of volunteers for AETs was also found to be limited: 34% of the patients who volunteered for an AET finally entered the trial [11]. The majority of the volunteers was excluded because of co morbid disorders. Investigators of the STAR*D trial [12] used less stringent inclusion criteria in order to obtain more generalizable results. In their study 22% of the included patients would have been eligible for AETs and had better treatment outcome than non-eligible patients [13]. However, the generalizability of STAR*D to routine clinical practice may still be limited, due to exclusion of prior non-responders to the study drugs, and the use of a minimum baseline severity. In addition, the generalizability of STAR*D to non-US health settings is unclear [14]. In the present study, we investigated in routine outpatient-care to what proportion of depressive patients the results of AETs would apply. We chose to limit the AETs to classical RCTs, since in national and international treatment guidelines [15-17], classical RCTs are considered the most robust evidence for efficacy. We do, however, expect that in the near future the results from more pragmatic trials like STAR*D and GENDEP [12,18] will influence guidelines. We applied the most frequently used inclusion and exclusion criteria for classical AETS to a large consecutive series of patients. Comprehensive data on patients' characteristics were available through the extensive Routine Outcome Monitoring (ROM) system. In addition, we investigated whether eligible patients differ from non-eligible patients in treatment outcome.

METHODS

The Dutch mental health care system and treatment steps for major depression

In the Netherlands, health insurance is obligatory for all citizens and regulated by the government. Mental health care is easily accessible and not restricted by the financial means of individual patients. The Dutch mental health care system is organized in a stepped-care-manner and uses evidence based treatment guidelines. Patients with mood complaints visit their general practitioner (GP) first. The treatment guidelines recommended that patients with mild to moderate depression should be treated with psychotherapy of pharmacotherapy, based on the patient's preferences [17]. Patients with severe depression should preferably start a pharmacotherapy. Rating of severity is based on clinical judgment. Reasons to refer patients to a regional mental health provider (RMHP) are a preference of patients for psychotherapy (not provided by GPs), more severe, recurrent or refractory depression or the presence of co morbid psychiatric or somatic disorders. After baseline assessment and a clinical interview at our RMHP, patients were offered treatment steps as recommended by the guidelines. Patients suffering from moderate to severe major depression could choose between psychotherapy and antidepressants. For severe depression antidepressants were the first choice. When patients were already on antidepressants the dose was optimized or patients were offered to switch to another antidepressant or start psychotherapy.

Routine Outcome Monitoring

In 2002, the RMHP Rivierduinen (service area with 1.1 million inhabitants), in collaboration with the University Medical Hospital Leiden, implemented ROM and evidence based, stepped care protocols. In ROM, all patients referred to the RMHP for treatment of a mood, anxiety or somatoform disorder have an extensive baseline assessment. Treatment progress is then assessed at three to four monthly intervals and before starting a new treatment step. The baseline assessment comprises a standardized diagnostic interview (Mini-International Neuropsychiatric Interview Plus [19]), the collection of sociodemographic and socioeconomic data, the administration of disease specific severity-scales, and general measures of health. For a more extensive description of ROM we refer to the design paper [20].

Patients

To examine the eligibility of depressive outpatients for AETS, we included all outpatients with a DSM-IV diagnosis of a current major depressive disorder as established by the Mini-International Neuropsychiatric Interview (MINIplus) [19], who sought treatment at the RMHP Rivierduinen from January 2002 until January 2007. The MINIplus does not yield a hierarchy in primary disorder and co morbid disorders. We included all patients with a major depressive disorder, regardless of the fact whether depression was the primary diagnosis

as determined by the treating clinician or a so-called co morbid disorder. We decided to do so since primary clinical diagnosis is a concept often used in clinical practice but not well defined in literature and is depending heavily on the personal expertise of the individual clinicians. Including patients based on primary clinical diagnosis of depressive disorder only would have led to selection bias and results in a less well defined sample. Since the presence of a primary clinical diagnosis of depression might influence treatment outcome, we controlled for it in the analyses on the influence of eligibility on treatment outcome.

In order to examine the influence of eligibility to AETs on treatment outcome, we selected all patients in our sample with at least one follow up assessment in ROM (follow-up group). The treatment outcome of the first treatment step was explored in this project. We examined possible selection bias by comparing the patient characteristics of the follow-up group and the lost-to-follow-up group. We conducted an extensive chart review in the follow-up group in order to obtain information on primary clinical diagnosis and treatment modality. In order to allow comparison with classical AETs, we defined treatment outcome in the same dichotomous variables used in AETs:

1. Proportion of responders: 50% reduction of the baseline score on the Montgomery Asberg Depression Rating Scale (MADRS) [21].
2. Proportion of remitters: MADRS≤10.

Commonly used exclusion criteria for Antidepressant Efficacy Trials

In an extensive review of the literature Zimmerman and co-workers identified exclusion criteria that where consistently used in AETs published between 1994 and 1999 in the top-five Impact Ranking journals in the US [9,22]. We expanded this search by inclusion of AETs published between 1994 and 2007, not only in the aforementioned journals, but also in the remaining journals of the top-ten Impact Ranking psychiatric journals of 2005. With our expanded search, we obtained 17 additional articles on AETs [1,23-36,37,38]. No additional exclusion criteria for AETs were identified. The commonly used exclusion criteria, identified by Zimmerman and co-workers [9] are listed below, together with the operationalisations for our sample.

1. ***History of DSM-IV manic or hypomanic episodes***
   At least one (hypo) manic episode on the MINIplus.
2. ***Experiencing psychotic features during the current episode of depression***
   Diagnosis of a current depression with psychotic features on the MINIplus.
3. ***Significant risk of suicide***
   In our sample, suicidality was assessed with the corresponding item on the MADRS item 8. Patients with a score of 3 or higher were considered to meet this exclusion-criterion.

The item is a Likert-scale from 0-6:

0 Enjoys life, takes it as it is.

2 Tired of life, only transient suicidal thoughts.

4 Probably better off dead. Suicidal thoughts often occur and suicide is considered to be a possible solution. No specific plans.

6 Explicit plans to commit suicide. Active preparation of suicide.

4. Alcohol or drug abuse or dependency within the last six months

Diagnosis of current abuse or dependence on drugs or alcohol on the MINIplus.

5. Mild depression, as determined by low baseline score on the Hamilton depression-scale

The most commonly used threshold for inclusion in an efficacy trial is a minimum score of 18 (HAMD 17 items) or a minimum score of 20 (HAMD 21 items) on the Hamilton depression scale [10,39]. Because in our setting the MADRS is used to asses depression severity, an equivalent of the HAMD score was computed using three previously developed regression equations based on three trials that compared the MADRS and the HAMD17 in outpatients Mittmann et al. [40] (A); Hawley et al., [41] (B) and Zimmerman et al. (C) [42]. Since the Item Response Theory (IRT) has recently been proven to be a probably more reliable method of conversion of the MADRS into the HAMD17 as well [43,44], we also used the IRT method to compute proportions of patients not meeting the criterion of minimum baseline severity.

1. (A) MADRS = 1.23 x HAMD - 0.30 (cutoff MADRS = 21.8)
2. (B) MADRS = 1.30 x HAMD + 0.7 (cutoff MADRS = 24.1)
3. (C) MADRS = 1.43 x HAMD + 0.87 (cutoff MADRS = 26.6)

6. Presence of underlying dysthymic disorder

Diagnosis of dysthymic disorder on the MINIplus.

7. Illness duration of less than 4 weeks or more than 2 years

Duration of less than 4 weeks or more than 2 years of the current episode is an exclusion criterion for antidepressant efficacy trials. Unfortunately, in our sample no reliable information on the duration of the current episode of the major depression was available. Therefore, we could not use this exclusion criterion in our analysis.

8. Presence of co morbid non-depressive, non-substance use Axis I disorders

Diagnosis of anxiety disorder, somatoform disorder, eating disorder, or attention deficit hyperactivity disorder on the MINIplus.

9. Presence of borderline personality disorder

In our setting, the Dimensional Assessment of Personality Pathology, short Dutch version DAPP-SF [45,46] was used as a screening instrument for personality-pathology. Stringent and less stringent cut-off scores were used to identify patients with a possible personality disorder within a population suffering from mood-, anxiety-, and somatoform disorders [46]. Quartiles (low score-intermediate low-intermediate high-high score) were computed for the patients in our sample on (weighted) scores for the dimensions Emotional Dysregulation, Dissocial Behavior and Inhibition. The scores were weighted by the factor loadings derived from research on psychometrics of the DAPP-SF [46]. In our sample, patients with a cut-off of 3.7 and a "high score" on all three dimensions were considered to meet the exclusion-criterion of borderline personality according to "*stringent criteria*". Patients with a cut-off of 2.6 and a "high score" on all three dimensions were considered to meet the exclusion-criterion of borderline personality disorder according to "*less stringent criteria*".

Statistical analysis

For each exclusion criterion, we determined the percentage of patients that met the criterion. For the DAPP-SF quartiles were computed for (weighted) scores. The scores were weighted by the factor loadings derived from research on psychometrics of the DAPP-SF [20]. In our sample, there were missing values for the MADRS (n=103) and the DAPP-SF (n= 415). Comparison of complete cases and cases with missing data showed differences on many variables. Therefore it is likely that the missing data were not missing-completely-at-random (MCAR). Complete case analysis is likely to yield biased estimates [47]. Therefore, the MICE (multivariate imputation by chained equations [48]) method was used to estimate missing values for MADRS. With these imputed data, we computed the percentage of patients meeting the exclusion criteria of *Mild Depression* and *Significant Risk of Suicide*. We did not impute missing values for the DAPP-SF, as this instrument consists of dimensional components that we considered too complex to predict. If the score for the DAPP-SF was missing for a patient, we considered the patient as *not meeting* the exclusion criterion of *Presence of Borderline Personality Disorder*. Comparison of proportion of responders and remitters in the eligible and non-eligible patient-groups were performed by Chi-square tests. The influence of the exclusion criteria and "eligibility for RCTs" on treatment outcome was computed by logistic regression after MICE. Odds-ratios (OR) and their confidence intervals were computed by using the robust standard error. Statistical analyses were performed with SPSS 16.0 and STATA10.0.

RESULTS

Patients

4157 outpatients were assessed at baseline between January 2002 and January 2007. Of these patients, 1653 suffered from a current major depressive disorder according to the MINIplus. The demographic features of the 1653 patients are described in table 1.

Table 1. Demographics.

N=1653	Percentage	Mean (+SD)
Age in years		38.19 (SD 11.68)
Gender	33.3% male; 66.7% female	
Marital situation Children living at home	37.9% married/living together 13.2% divorced/widowed 25.5 single 23.4% unknown 31.9% yes 43.4% no 23.7% unknown	
Professional situation	16.5% unemployed 27.8% employed 0.7% retired 26.8% sickness/disability benefit 28.2% unknown	
Education	9.2% primary school or less 25.4% secondary school, lower level 29.5% secondary school intermediate/high level 12.1% academic or higher professional education 23.8% unknown	
Ethnicity	64.5% born in the Netherlands 4.1 % born in Morocco/Turkey 2.1% born in Suriname/Antilles 5.6% born elsewhere 23.4% unknown	
Ethnicity II	60.0% parent(s) born in the Netherlands 4.6% parent(s) born in Morocco/Turkey 2.3% parent(s) born in Suriname/Antilles 8.8% parent(s) born elsewhere 23.4% unknown	
MADRS at baseline		26.76 (SD 7.52)

SD = standard deviation

Application of commonly used exclusion criteria for AETs

Bipolarity and Psychotic Features

A total of 25 of the 1653 patients (1.5%) had at least one (hypo) manic episode (current or history). 31 patients suffered from a depression with psychotic features (1.9%). There was no overlap between these two groups. Following the approach by Zimmerman and colleagues [9,49], we excluded these 56 patients (3.4%) from further analysis. The other exclusion criteria were examined on the remaining 1597 patients.

Suicidality

Of the 1597 patients 241 patients (15.1%) would have been excluded from AETs because of suicide risk.

Alcohol or drug abuse/dependence

142 of the 1597 patients (8.9%) met the exclusion-criterion of current abuse/dependence on drugs or alcohol.

Severity of the depression at baseline

According to the first regression equation (A), 435 of the 1597 patients (27.2%) did not meet the cut-off score of 18 on the HAMD17. The second and the third regression (B, C) equations yielded identical scores, and 664 of the 1597 patients (41.6%) had a score lower than 18 on the HAMD17. The IRT yielded almost identical proportions: 38.7 % (cut-off MADRS 24) – 44.5% (cut-off MADRS 25) of the patients had a score lower than 18 on the HAMD17.

Co morbid Dysthymic Disorder

136 of the 1597 patients (8.5%) met the exclusion-criterion for a co morbid dysthymic disorder.

Other co morbid Axis I disorders

1003 of the 1597 patients (62.8 %) had co morbid diagnoses on Axis I according to the MINIplus. 730 patients (45.7%) had at least one anxiety disorder.180 patients (11.3%) had at least one somatoform disorder. Another 32 patients (2.0%) had other co morbid disorders.

Personality Pathology

31.6–61.6% of the 1597 patients in our sample may have had some form of personality pathology according to the DAPP-SF. Within this group, the estimated percentage of patients suffering from a borderline personality disorder ranges from 3 patients (0.2%, stringent criteria) to 112 patients (7.0%, less stringent criteria).

Percentage of patients eligible for Antidepressant Efficacy Trials and comparison with previous research

Finally, the sample of 1653 depressed outpatients was filtered by stepwise application of the exclusion criteria. Only 17.0%–24.5% of our patients would have been eligible for AETs. Stepwise application of the exclusion criteria is described in figure 1. Comparison of the incidence of the individual exclusion criteria in our sample with previous research [9] is described in table 2.

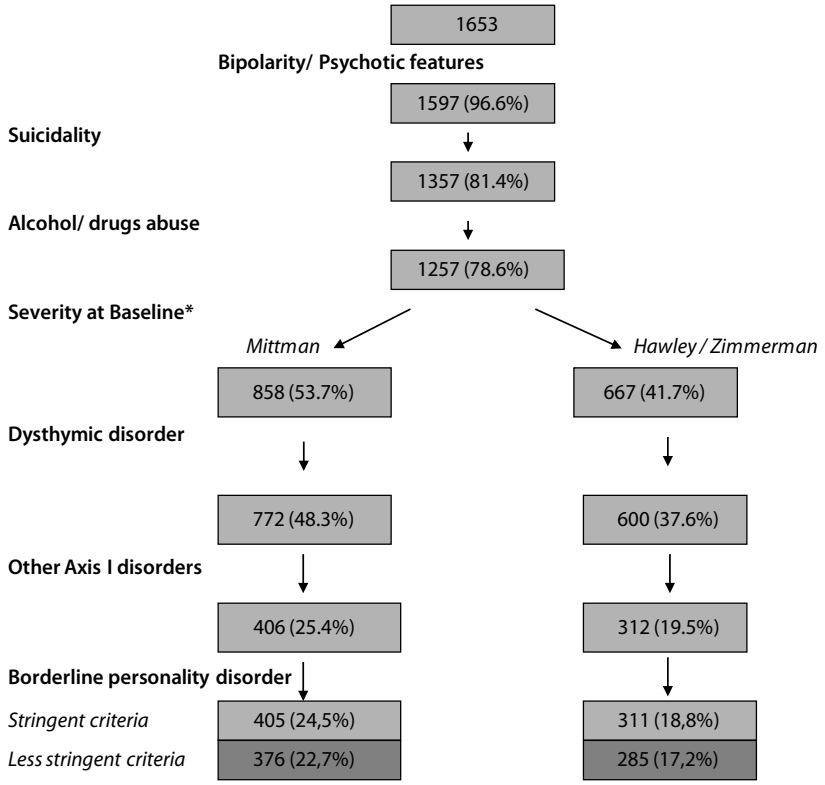

Figure 1. Stepwise Application of Commonly Used Exclusion Criteria and the Resulting Percentages of Patients Eligible for Antidepressant Efficacy Trials.

* Severity at baseline was assessed with the Montgomery Asberg Rating Scale for Depression. Equivalent Hamilton rating Scale for Depression scores (17 items version) were calculated using three previously developed regression equations.

Table 2. Comparison of incidences of exclusion criteria (%) between our sample and Zimmerman's sample [9].

	Current research Percentages of excluded patients	**Previous research Percentages of excluded patients**
Bipolarity/ psychotic features	3.4	15.3
Suicidality	15.2	19.8
Alcohol/drugs	8.6	7.8
Severity at baseline	27.2–41.6	54.3
Dysthymic disorder	8.5	8.9
Other Axis I disorders	62.8	68.3
Borderline personality pathology	0.2–7.0	11.9

Follow-up group

From the 1653 patients suffering from major depression, 46% (n=774) had a follow-up assessment. Extensive chart-review was done for those 774 patients. 148 patients had to be excluded from further follow-up analysis due to suspected bipolarity/psychotic features, admission to an inpatient-clinic during follow-up, remission on the MADRS at baseline or a time-span between baseline and follow-up assessment which we considered either to be too short or too long to provide reliable information. Finally, 626 patients were selected for follow-up analysis. In 4% of the 626 patients, information on primary clinical diagnosis for was missing. Patient selection is described in figure 2.

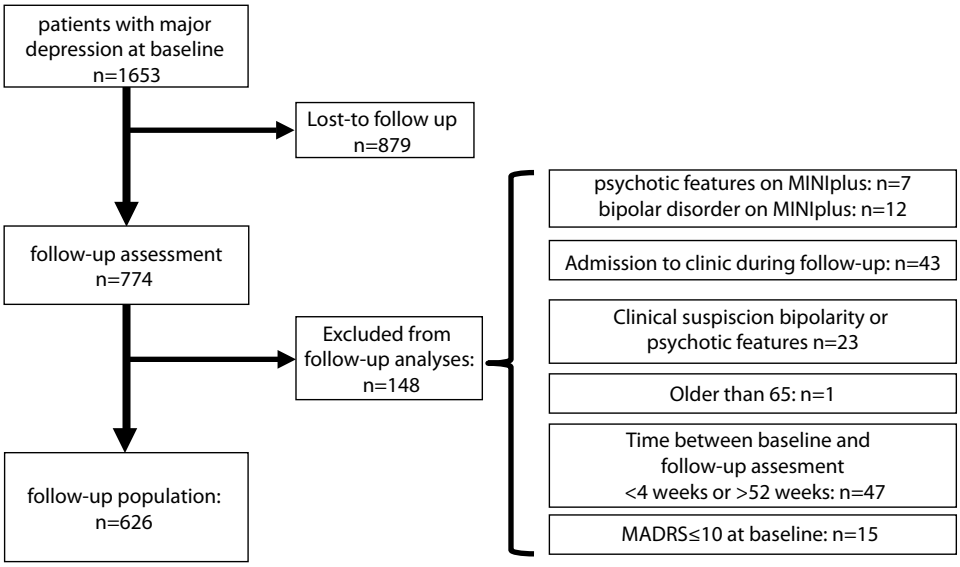

Figure 2. Selection of the follow-up group.

In chart review, we identified that 54% of the selected patients in the follow-up group received antidepressants, either as solo treatment or in combination with other treatment modalities. Five treatment modalities were identified: "antidepressants (AD)" (13%), "individual psychotherapy (IP)" (27%, mostly cognitive behavioral therapy or interpersonal therapy), "combination of antidepressants and individual psychotherapy (AD+IP)" (27%), "antidepressants and social supportive therapy (AD+SST)" (14%) and "other treatment/ insufficient information" (19%). The mean time-span between start of treatment and follow-up assessment was as follows: AD 20.8 weeks (CI 18.7–22.9); IP 20.1 weeks (CI 18.5–21.6); AD+IP 21.5 weeks (CI 20.0–23.1); AD+SST 21.6 weeks (CI 19.9-23.3); other 19.1 weeks

(CI 17.1–21.1). In 113 patients treatment was primarily started for a clinical diagnosis other than major depression, of whom 88 patients received psychotherapeutic treatment focussed specifically on anxiety/somatoform disorders.

Lost to follow-up analysis

The follow-up group and the lost-to-follow-up group did not differ on most exclusion criteria. The follow-up group only differed from the lost-to-follow-up group in larger proportions of patients with a generalized anxiety disorder (7.3% vs. 4.6%, X^2=5.08, df1, p=0.02) and depression with psychotic features (0.9% vs. 2.7%, X^2=7.42, df1, p=0.01). Based on these results, selection bias was considered to be very small.

Influence of eligibility on treatment outcome

In the follow-up group, 28% of the patients met the criteria for response and 21% of the patients met the criteria for remission. There were no significant differences in response-percentages between the patients who would have been eligible for AETs (25%) and those who were not (28%), X^2=0.26, df1, p=0.61. Remission percentages did not differ either: 16% (eligible patients) vs. 23 % (non-eligible patients), X^2=1.80, df1, p=0.18. The influence of patient features commonly used as exclusion criteria on response and remission was examined in multivariate regression models. The following variables were entered as covariates in a multivariate regression model: risk of suicide; minimum baseline severity of depression; co morbid substance dependency/abuse; co morbid dysthymia, co morbid anxiety disorder, co morbid somatoform disorder, other co morbid Axis I disorders. "Primary clinical diagnosis" and "treatment modality" were entered in the model as possible confounders. Overall, the explained variance (R-square) was very low for remission (4.1%) and response (1.4%). Only "the criterion of minimum baseline severity" contributed to remission (OR 2.0, CI 1.3–3.1). None of the exclusion criteria contributed significantly to response. The influence of "eligibility for AETs", which we defined as "not meeting any of the exclusion criteria", was investigated in a separate model and did not contribute significantly to response (OR 0.90, CI 0.5–1.8) nor remission (OR 1.0, CI 0.5–2.0).

DISCUSSION

We evaluated the eligibility for inclusion in AETs in 1653 outpatients with a major depressive disorder in a Dutch general psychiatric outpatient setting. We followed a model developed for the consistency of exclusion criteria used in AETs [9,22]. We found that the majority of patients in our sample (75%) did not meet the inclusion criteria. The most common criteria for inclusion that would not have been met were "minimum baseline severity of 18 on the Hamilton rating scale" and "no co morbid Axis I disorder". In addition, we examined the

influence of eligibility on treatment outcome. The influence of exclusion criteria on response and remission appears to be small. Only the exclusion of mild depression contributed to improvement of treatment outcome in our sample. Exclusion of less severely depressed patients from the analyses resulted in better treatment outcome. Milder depression is highly prevalent in daily practice and more research into treatment effectiveness in milder depression is warranted.

Comparison with previous research: eligibility for AETs

Our findings are in line with those of previous research [8,11,13,49]. The percentage of patients eligible for participation in AETs in our study was higher than in earlier research but similar to the latest report on eligibility in the STAR*D trial [13]. An explanation for the larger proportion of eligible patients in our sample might be the fact that the percentage of patients meeting the criterion of minimum baseline severity was larger in our sample. This might be due to the way in which the Dutch health care system is organized. First, there is no (financially) limited access to mental health care. Poor socioeconomic status has been shown to be associated with more severe pathology and co morbidity [50]. Therefore, we expected a higher percentage of patients with more severe depression in our sample. We also expected higher percentages of co morbid Axis I disorders, but the prevalence of co morbidity was similar to previous research. Another explanation for the higher percentage of patients that met the criteria for baseline severity is the role of the GP as 'gate keeper' in Dutch health care. Still, a considerable part of our patients did not meet the criteria for minimum baseline severity (27–42%).We found lower percentages of bipolarity / psychotic features and borderline personality disorders in our sample. In our RMHP those patients are often directly (preceding ROM baseline assessment) referred to specialized teams, which might explain the low prevalence in our sample.

Comparison with previous research: influence of eligibility on treatment outcome

In contrast to the recent STAR*D report [13], we found no differences in treatment outcome between eligible and non-eligible patients. Together with the marginal explained variance that we found in our model, this suggests that other patient features are more associated with treatment outcome than eligibility for AETs. Many patients, either eligible or not, would not be willing or able to participate in AETs. Participants might differ from non-participants in: sociodemographic/socioeconomic status, motivation/adherence to treatment and the interaction between clinician and patients. This might also partially explain the differences between our results and the ones found in the STAR*D report [13]. In the STAR*D trial, much effort has been undertaken to improve adherence to treatment and to motivate the participating patients and clinicians [12]. It is possible that by "controlling" for these aspects an association between eligibility and treatment outcome can be detected. Unfortunately,

the magnitude of the influence of eligibility on treatment outcome is not reported in the STAR*D report and therefore not available for comparison.
Treatment outcome in our study was less favorable than the outcomes typically found in classical AETs and also less favorable than the outcome in STAR*D. A thorough comparison and exploration of differences between the outcomes in RCTs, in more pragmatic trials like STAR*D, and in our ROM project will be important for daily practice. We are currently performing such a comparison and exploration.

Strengths

The use of Routine Outcome Monitoring in daily mental health care provided comprehensive data on a large number of patients As the only restriction for participation is sufficient language competence and ability to complete computerized or written questionnaires, the results of this type of data collection are very representative of and generalizable to 'real-life' psychiatric practice. Furthermore, we consider the fact that the Dutch health care system provides unrestricted access to mental health care as a strong quality of this research. It diminishes the possibility of selection bias even further.

Limitations

There was a considerable loss to follow-up in our study. In 22% of the lost to follow-up, patients dropped out of treatment directly after baseline assessment and in 38% of the lost to follow-up, patients stayed in treatment, but we had no information on their treatment course. The major reasons for drop-out are unclear; patients might have recovered, were perhaps unsatisfied with the offered treatment or treatment results, or had poor compliance. As 38% of the lost-to-follow-up patients remained in treatment, loss to follow-up may also have resulted from factors hampering the ROM follow-up assessments, such as administrative issues or a reduced adherence of clinicians to the ROM methodology. A large loss to follow up might be a problem in all studies with a more naturalistic design. For example, STAR*D had reached a loss-to-follow-up of 48% in step II of the study. Of the 4790 patients who were screened at baseline, 12% was not willing to participate; 3% did not meet inclusion criteria; 8% had an HAMD <14 or no data on the HAMD; and 25% left the study [12]. Although we had a considerable loss to follow-up, the follow-up group was very similar to the lost-to-follow-up group with respect to criteria for eligibility. We therefore expect the influence of the loss to follow-up on our results to be small.

The absence of information on illness duration is another limitation of this study. Although we expect not to have included patients with illness duration shorter than four weeks as most patients are seen several times by their GP before referral, it is however possible that patients were depressed for more than 2 years. This might have lead to an overestimation of the amount of eligible patients. Furthermore, a possible suboptimal diagnostic assessment of borderline personality disorder the fact that we had no information on physical health

problems (not included in Zimmerman's model on exclusion criteria, but still often used as an exclusion criterion) might also have led to an overestimation of the amount of eligible patients. On the other hand, there might be some underestimation of the eligibility in our sample, due to the fact that no data were available on patients who were too ill to complete questionnaires. Not all the patients in our sample were treated with antidepressants. A considerable proportion received other treatment (i.e. psychotherapy) for their depression. However, the percentage of eligible patients turned out to be equal in the antidepressants-group and the other-treatment-group. For comparability with former research, we used the model of Zimmerman et al. which does not take differences between AETs, like active versus placebo controlled, into account. Differences in AET architecture will probably influence eligibility, but were not investigated in the present study. Finally, to optimize comparability in treatment outcome with classical RCTs, we used the same definitions of outcome as RCTs: response and remission, determined by a cut-off score. This dichotomization of scales might lead to loss of information compared to continuous outcomes [18] .

REFERENCE LIST

1. Fava GA, Ruini C, Rafanelli C: Sequential treatment of mood and anxiety disorders. *J Clin Psychiatry* 2005, 66: 1392-1400.
2. IJff MA, Huijbregts KM, van Marwijk HW, Beekman AT, Hakkaart-van Roijen L, Rutten FF *et al*.: Cost-effectiveness of collaborative care including PST and an antidepressant treatment algorithm for the treatment of major depressive disorder in primary care; a randomized clinical trial. *BMC Health Serv Res* 2007, 7: 34.
3. Stewart JW, McGrath PJ, Quitkin FM: Can mildly depressed outpatients with atypical depression benefit from antidepressants? *Am J Psychiatry* 1992, 149: 615-619.
4. Mulder RT, Frampton C, Joyce PR, Porter R: Randomized controlled trials in psychiatry. Part II: their relationship to clinical practice. *Aust N Z J Psychiatry* 2003, 37: 265-269.
5. Tunis SR, Stryer DB, Clancy CM: Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003, 290: 1624-1632.
6. Wells KB: Treatment research at the crossroads: the scientific interface of clinical trials and effectiveness research. *Am J Psychiatry* 1999, 156: 5-10.
7. Licht RW, Gouliaev G, Vestergaard P, Frydenberg M: Generalizability of results from randomized drug trials. A trial on antimanic treatment. *Br J Psychiatry* 1997, 170:264-7.
8. Zetin M, Hoepner CT: Relevance of exclusion criteria in antidepressant clinical trials: a replication study. *J Clin Psychopharmacol* 2007, 27: 295-301.
9. Zimmerman M, Chelminski I, Posternak MA: Exclusion criteria used in antidepressant efficacy trials: consistency across studies and representativeness of samples included. *J Nerv Ment Dis* 2004, 192: 87-94.
10. Zimmerman M, Posternak MA, Chelminski I: Symptom severity and exclusion from antidepressant efficacy trials. *J Clin Psychopharmacol* 2002, 22: 610-614.
11. Partonen T, Sihvo S, Lonnqvist JK: Patients excluded from an antidepressant efficacy trial. *J Clin Psychiatry* 1996, 57: 572-575.
12. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA *et al*.: Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials* 2004, 25: 119-142.
13. Wisniewski SR, Rush AJ, Nierenberg AA, Gaynes BN, Warden D, Luther JF *et al*.: Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR*D report. *Am J Psychiatry* 2009, 166: 599-607.
14. Hatcher S: The STAR*D trial: the 300 lb gorilla is in the room, but does it block all the light? *Evid Based Ment Health* 2008, 11: 97-99.
15. Karasu T, Gelenberg AJ, Merriam A, Wang P. Practice guidelines for the treatment of patients with major depressive disorder. Second Edition. 2000. American Psychiatric Association.
16. Anderson I, Pilling S, Barnes A, Bayliss L, Bird V. The NICE guideline on the treatment and management of depression in adults. Edited by National Collaborating Centre for Mental Healt, National Institute for Health and Clinical Excellence. Updated version 2010. 1-1-2009. London:The British Psychological Society & The Royal College of Psychiatrists.
17. Trimbos Institute. Richtlijnherziening van de Multidisciplinaire richtlijn Depressie (eerste revisie) . 2009.
18. Uher R, Maier W, Hauser J, Marusic A, Schmael C, Mors O *et al*.: Differential efficacy of escitalopram and nortriptyline on dimensional measures of depression. *Br J Psychiatry* 2009, 194: 252-259.
19. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E *et al*.: The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998, 59 Suppl 20: 22-33.

20. de BE, den Hollander-Gijsman ME, van Rood YR, van der Wee NJ, Giltay EJ, van Noorden MS *et al*.: Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clin Psychol Psychother* 2010.
21. Asberg M, Montgomery SA, Perris C, Schalling D, Sedvall G: A comprehensive psychopathological rating scale. *Acta Psychiatr Scand Suppl* 1978, 5-27.
22. Posternak MA, Zimmerman M, Keitner GI, Miller IW: A reevaluation of the exclusion criteria used in antidepressant efficacy trials. *Am J Psychiatry* 2002, 159: 191-200.
23. Benkert O, Szegedi A, Philipp M, Kohnen R, Heinrich C, Heukels A *et al*.: Mirtazapine orally disintegrating tablets versus venlafaxine extended release: a double-blind, randomized multicenter trial comparing the onset of antidepressant response in patients with major depressive disorder. *J Clin Psychopharmacol* 2006, 26: 75-78.
24. Bielski RJ, Ventura D, Chang CC: A double-blind comparison of escitalopram and venlafaxine extended release in the treatment of major depressive disorder. *J Clin Psychiatry* 2004, 65: 1190-1196.
25. DeMartinis NA, Schweizer E, Rickels K: An open-label trial of nefazodone in high co morbidity panic disorder. *J Clin Psychiatry* 1996, 57: 245-248.
26. Derubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM *et al*.: Cognitive therapy vs medications in the treatment of moderate to severe depression. *Arch Gen Psychiatry* 2005, 62: 409-416.
27. Detke MJ, Lu Y, Goldstein DJ, Hayes JR, Demitrack MA: Duloxetine, 60 mg once daily, for major depressive disorder: a randomized double-blind placebo-controlled trial. *J Clin Psychiatry* 2002, 63: 308-315.
28. Dinan TG: Efficacy and safety of weekly treatment with enteric-coated fluoxetine in patients with major depressive disorder. *J Clin Psychiatry* 2001, 62 Suppl 22: 48-52.
29. Golden RN, Nemeroff CB, McSorley P, Pitts CD, Dube EM: Efficacy and tolerability of controlled-release and immediate-release paroxetine in the treatment of depression. *J Clin Psychiatry* 2002, 63: 577-584.
30. Goldstein DJ, Mallinckrodt C, Lu Y, Demitrack MA: Duloxetine in the treatment of major depressive disorder: a double-blind clinical trial. *J Clin Psychiatry* 2002, 63: 225-231.
31. Goldstein DJ, Lu Y, Detke MJ, Wiltse C, Mallinckrodt C, Demitrack MA: Duloxetine in the treatment of depression: a double-blind placebo-controlled comparison with paroxetine. *J Clin Psychopharmacol* 2004, 24: 389-399.
32. Langworth S, Bodlund O, Agren H: Efficacy and tolerability of reboxetine compared with citalopram: a double-blind study in patients with major depressive disorder. *J Clin Psychopharmacol* 2006, 26: 121-127.
33. Mulder RT, Joyce PR, Frampton CM, Luty SE, Sullivan PF: Six months of treatment for depression: outcome and predictors of the course of illness. *Am J Psychiatry* 2006, 163: 95-100.
34. Shelton RC, Haman KL, Rapaport MH, Kiev A, Smith WT, Hirschfeld RM *et al*.: A randomized, double-blind, active-control study of sertraline versus venlafaxine XR in major depressive disorder. *J Clin Psychiatry* 2006, 67: 1674-1681.
35. Sir A, D'Souza RF, Uguz S, George T, Vahip S, Hopwood M *et al*.: Randomized trial of sertraline versus venlafaxine XR in major depression: efficacy and discontinuation symptoms. *J Clin Psychiatry* 2005, 66: 1312-1320.
36. Trivedi MH, Pigotti TA, Perera P, Dillingham KE, Carfagno ML, Pitts CD: Effectiveness of low doses of paroxetine controlled release in the treatment of major depressive disorder. *J Clin Psychiatry* 2004, 65: 1356-1364.
37. Fabre LF, Abuzzahab FS, Amin M, Claghorn JL, Mendels J, Petrie WM *et al*.: Sertraline safety and efficacy in major depression: a double-blind fixed-dose comparison with placebo. *Biol Psychiatry* 1995, 38: 592-602.
38. Stahl SM: Placebo-controlled comparison of the selective serotonin reuptake inhibitors citalopram and sertraline. *Biol Psychiatry* 2000, 48: 894-901.

39. Hamilton M: Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967, 6: 278-296.
40. Mittmann N, Mitter S, Borden EK, Herrmann N, Naranjo CA, Shear NH: Montgomery-Asberg severity gradations. *Am J Psychiatry* 1997, 154: 1320-1321.
41. Hawley CJ: Depression rating scales can be related to each other by simple equations. 1998.
42. Zimmerman M, Posternak MA, Chelminski I: Derivation of a definition of remission on the Montgomery-Asberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. *J Psychiatr Res* 2004, 38: 577-582.
43. Carmody TJ, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D *et al*.: The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. *Eur Neuropsychopharmacol* 2006, 16: 601-611.
44. Uher R, Farmer A, Maier W, Rietschel M, Hauser J, Marusic A *et al*.: Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychol Med* 2008, 38: 289-300.
45. Livesley WJ: The Dimensional Assessment of Personality Pathology (DAPP) Approach to Personality Disorder. 2006.
46. van Kampen D, de Beurs E, Andrea H: A short form of the Dimensional Assessment of Personality Pathology-Basic Questionnaire (DAPP-BQ): the DAPP-SF. *Psychiatry Res* 2008, 160: 115-128.
47. Donders AR, van der Heijden GJ, Stijnen T, Moons KG: Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006, 59: 1087-1091.
48. Royston P: Multiple imputation of missing values: update. *Stata Journal* 2005, 5: 188-201.
49. Zimmerman M, Mattia JI, Posternak MA: Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *Am J Psychiatry* 2002, 159: 469-473.
50. Lesser IM, Leuchter AF, Trivedi MH, Davis LL, Wisniewski SR, Balasubramani GK *et al*.: Characteristics of insured and noninsured outpatients with depression in STAR(*)D. *Psychiatr Serv* 2005, 56: 995-1004.