Universiteit Leiden
The Netherlands

Are depression trials generalizable to clinical practice? Something clinicians always wanted to know about RCTs, but were afraid to ask.....

Lem, R. van der

Citation
Lem, R. van der. (2013, June 12). *Are depression trials generalizable to clinical practice? Something clinicians always wanted to know about RCTs, but were afraid to ask*.. Retrieved from https://hdl.handle.net/1887/20950

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/20950

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/20950 holds various files of this Leiden University dissertation.

Author: Lem, Rosalind van der
Title: Are depression trials generalizable to clinical practice? Something clinicians always wanted to know about RCTs, but were afraid to ask.....
Issue Date: 2013-06-12

Chapter 7

Summary and General Discussion

SUMMARY OF OUR FINDINGS

In today's psychiatric practice in Western societies, most mental health care institutions have implemented treatment algorithms or guidelines for the treatment of major depressive disorder (MDD). These treatment algorithms/guidelines are all based on results from randomized clinical trials (RCTs), also called efficacy trials. In daily psychiatric practice, many clinicians have the impression that results found in trials (efficacy) are better than the results of the same therapies in routine care (effectiveness). In this thesis, we investigated whether the clinicians' impression that efficacy is higher than effectiveness is correct and, if the impression is substantiated, which factors explain the difference. Criticism about the generalizability of results from RCTs to daily practice has often been heard. Clinicians believe that the possible difference in effect is explained by differences between their patients and participants in MDD trials. Clinicians are supported by previous research on the generalizability of results from MDD trials. It has been shown that only a minority of "real life" patients is eligible for participation in RCTs, because of the stringent criteria for patient selection [1-3] and perhaps also because of (un)intended selection due to the methodology of recruitment of participants in trials [4]. The STAR*D trial [5] found that participants who were eligible for "classical" MDD trials had a beneficial outcome compared to participants who were not [6]. The STAR*D trial, with very broad inclusion criteria, has many similarities with daily practice. However STAR*D also has characteristics of an RCT, like the use of a baseline severity threshold for inclusion, no possibilities for patient preferences in the first treatment step, and a large investment in treatment adherence of both therapists and participants. Despite the broadness of the inclusion criteria, it is possible that the RCT characteristics of STAR*D may still have limited its generalizability to daily practice. For this project we derived our data directly from daily practice through Routine Outcome Monitoring (ROM). By doing so, we were able to investigate whether clinicians are right when they state that treatment outcome in daily practice is less hopeful than in efficacy trials. Subsequently, we were curious to see whether the evidence from the STAR*D trial of a better treatment outcome in "RCT-eligible" patients could be replicated. If clinicians would, hypothetically, exclude all of their non-eligible patients, would their treatment results improve?

In order to assess whether effectiveness is really lower than efficacy (chapter 2), we compared the within group efficacy reported in fifteen meta analyses on three types of MDD treatment; antidepressants, individual psychotherapy and a combination of both, with the effectiveness of the same treatments in daily practice, measured by ROM. A meta analysis provides an aggregated estimate of results found in RCTs. Meta analyses of RCTs are most often carried out to investigate whether the active drug/psychotherapy is superior to placebo (which is called the *between group* efficacy). However, we were not interested in this relative effect of active drug/psychotherapy, but in their overall or absolute effect (which is called the *within-group* efficacy). We compared this overall efficacy with the effectiveness

in "reality". Our overall conclusion in chapter 2 is that the impression of clinicians, that treatments in "reality" are not as effective as in scientific research, is true:

> ***Effectiveness of MDD treatment in daily practice is lower than efficacy results from RCTs on MDD treatment. This is the case for antidepressant treatment, individual psychotherapy as well as combination treatment.***

Above we mentioned that clinicians attribute the smaller treatment effects in "real life" to the fact that only a selection of patients is allowed to participate in RCTs. To investigate this, we first made an inventory of the exclusion criteria used in RCTs. Next we studied how many patients would have to be excluded if these criteria were applied in clinical practice, and then we compared treatment effectiveness in "real life" patients who meet the selection criteria versus patients who do not.

For inclusion in an MDD efficacy trial, in antidepressant efficacy trials (AETs) as well as in psychotherapy efficacy trials (PETs), participants indeed have to meet a set of eligibility criteria (in and exclusion criteria). These eligibility criteria are necessary to optimize the internal validity of the trial. In AETs, there is consistency in the use of exclusion criteria [7,8]. The most commonly used exclusion criteria in AETs are: not meeting a baseline severity threshold of 18 on HAMD17 [9]; co morbid Axis I disorders; co morbid Axis II disorders (in particular borderline personality disorder); suicidality and co morbid substance abuse. In the literature it is reported that only a minority of MDD patients from fee-for-service practices are eligible for AETs [1,2]. In this thesis, we investigated whether in the Netherlands in routine care also only a minority of patients with MDD would be eligible for AETs. We planned to do the same for PETs, yet studies on patient selection in PETs were absent. Therefore, we first had to investigate which exclusion criteria were used in a large set of PETs (chapter 4). We found that the following exclusion criteria were frequently used in PETs: not meeting a baseline severity threshold of 14 on HAMD17 [9]; co morbid substance abuse and antidepressant treatment prior to participation.

The next step was to apply the exclusion criteria of AETs (chapter 3) and PETs (chapter 4) to a "real life" (ROM) population (in which no selection takes place besides sufficient mastery of the Dutch language to complete the ROM questionnaires). We used a large dataset of MDD patients who sought treatment in Rivierduinen, a large regional mental health provider (RMHP). We found that clinicians are right when they state that their patients are very different from RCT participants.

- ***"Real life" MDD patients often do not meet the baseline severity threshold (42% for AET threshold and 22% for PET threshold).***

- ***"Real life" MDD patients do report suicidality (15%).***
- ***"Real life" MDD patients do have Axis I (63%) and Axis II co morbidity (7% borderline personality disorder).***
- ***"Real life" MDD patients do have co morbid substance abuse (9%).***
- ***"Real life" MDD patients do have often used antidepressants prior to referral to psychotherapeutic treatment (44%).***

Apart from selection based on explicit criteria also implicit selection may be important in RCTs, for instance with respect to sociodemographic and socioeconomic (SES) features. In chapter 5 we studied which SES features were reported in AETs and PETs. It became clear that educational level, socioeconomic status and income were reported insufficiently. However, for some features (age, gender, ethnicity, marital and employment status) enough data were available to enable comparison with "real life" patients (chapter 6). Our most striking finding was:

"Real life" MDD patients significantly less often have a paid job at time of treatment than RCT participants.

Having identified criteria that play a role in the selection of patients for RCTs and having demonstrated that application of these criteria in "real life" would indeed exclude a large group of patients from treatment, the next question is whether "real life" patients who are eligible for RCTs are doing better in treatment than "real life" patients who would be excluded. We found that exclusion of patients with mild depression, patients who used antidepressants prior to psychotherapy or patients without a paid job, improved treatment outcome in the remaining patient group, but only in a modest way. Besides, Axis I and Axis II co morbid disorders, substance abuse and suicidality were not associated with treatment outcome in our MDD patients. Furthermore the extent to which the difference in treatment outcome between RCTs and "real life" can be attributed to patient selection based on exclusion criteria is very small (explained variances 1–4% for the AET criteria; 4–11% for the PET criteria, dependent on definition of outcome). The same accounts for implicit selection based on the SES features age, gender, ethnicity, marital and employment status (explained variance 3–7%).

Therefore, our most striking overall finding was that:

In our "real life" patients, being eligible (meeting all criteria) for RCTs was not associated with a better treatment outcome.

GENERAL DISCUSSION

We found that if only RCT eligible patients were treated in daily practice and non eligible patients would be excluded, the treatment success in daily practice would not improve. So….yes, clinicians àre right that their MDD treatment results are less favorable than those from efficacy trials, and yes, they àre right that their patients differ very much from RCT participants due to the use of stringent exclusion criteria and (un)intended sociodemographic/socioeconomic patient selection by recruitment procedures. However…. the use of exclusion criteria and the selection of patients with a different socioeconomic status in RCTs do not explain the difference between efficacy and effectiveness. So, it might be that the items from patient selection that we analyzed are not the major threat to the generalizability of the results from MDD trials to daily practice as has been suggested in the past. In the next paragraphs, we will elaborate on the implications of our findings for clinical practice and the scientific field. We will seek further explanations for the difference between efficacy and effectiveness of MDD treatment. Although the effect of the use of exclusion criteria was modest on treatment outcome, we will however comment on the implications of our findings that patients suffering from minor depression, as well as patients who used antidepressants prior to their psychotherapy seem to benefit less from their treatment. We will also discuss the implications of our finding that patients without a paid job have a less favorable treatment outcome. Finally, we will discuss the limitations of our project and will conclude with recommendations both for future research and clinical practice.

Why do efficacy and effectiveness differ?

In this thesis, we found evidence for the assumption of clinicians that treatment results in daily practice are disappointing compared to those in MDD trials (of the same therapies). Of MDD patients in antidepressant trials, 34–47% reaches remission, whereas in daily practice only 21% of the patients are that fortunate after the first treatment step. For individual psychotherapy (cognitive behavioral therapy or interpersonal therapy), 34–58% of trial participants reach remission, while in daily practice only 27% of the patients reach remission after the first treatment step. Patients who receive combination therapy in daily practice reach remission in 21% of the cases, while in trials 45-63% do. We have shown that (un) intentional patient selection based on exclusion criteria or on socioeconomic grounds does not explain the difference between efficacy and effectiveness. Then, what does*? Does Dr. X, introduced in the Introduction section of this thesis, turn out to be a lousy therapist? Is the faith of younger colleagues in his knowledge and experience misplaced? Or do we have to look for other explanations for the difference between efficacy and effectiveness?*

Dr. X, now getting worried, provokingly states that the disappointing outcome results that we found in this thesis are typical for the RMHP Rivierduinen, for the Leiden area, or for Dutch psychiatry. Is he right? The effectiveness that we found is in line with the results of STAR*D

[5,10,11], which suggests that the modest treatment results are not typically from the RMHP Rivierduinen or Dutch psychiatric practice. The similarity of our results to those from STAR*D is a notable finding. As mentioned at the start of this chapter, STAR*D is a pragmatic trial, with on the one hand methodological characteristics of RCTs, but on the other hand resemblances with routine psychiatric practice. Much effort was put in treatment adherence and motivation, both at the side of patients and of clinicians. This may have inflated the treatment success. On the other hand, following patient's or doctor's preferences for a specific drug or psychotherapy, as is usual in daily clinical practice, was not allowed in the first treatment step in STAR*D. As the allowance of preference is associated with better treatment outcome [12-15], this might have diminished treatment success in STAR*D. In our population, no special effort was made to improve treatment adherence besides care as usual. Therefore, based on treatment adherence alone, one would probably expect less favorable treatment outcome in the ROM population than in STAR*D. On the other hand, in our daily practice population patient and doctor's preferences were of course allowed, which may have raised our treatment effect compared to STAR*D. These two factors together may have contributed to similar results in STAR*D and our ROM data. More emphasis on the improvement of treatment adherence, as is done in RCTs as well as in STAR*D, may improve the treatment results of daily practice. Furthermore, many other factors may contribute to the differences between efficacy and effectiveness. They may be features of patients, therapists, setting or RCT methodology. We will discuss them one by one in the following paragraph.

Patient features

Today, Dr. X's first patient is Ms. Y. Ms. Y is a moderately severe depressed and traumatized single, middle-aged woman who just lost her job and whose cat just died. Counseling sessions in a private practice and antidepressant treatment by her general practitioner did not improve her mood. She is somewhat sceptical about her referral to dr. X but is determined to give it a try and tell dr. X about all her problems in the first session. Right before his busy clinic starts, Dr. X quickly opens his mailbox. In his mailbox is an enthusiastic letter from a young colleague working in an academic center who asks psychiatrists to send in patients for a promising trial with a specific drug. What are the chances that Ms. Y will be willing to participate in this trial?

Likely, there are differences between RCT participants and daily practice patients, which we did not explore. For instance, participants in RCTs probably are a subgroup with a special motivation: they are willing to take the risk to be treated with a placebo. It is yet unknown which other specific characteristics this subgroup has and whether these characteristics may contribute to treatment outcome. Recruitment procedures might introduce (un) intentional selection bias by recruiting patients with a prognosis that differs from the "real life" population. Clinicians might not send patients with a poor prognosis for participation

to trials or these patients might not be motivated to participate. More importantly, in the Netherlands, and in other countries with a stepped health care system where the general practitioners (GPs) have the function of gatekeeper, many MDD patients with a good prognosis will be (successfully) treated by their GP or in private practice (so called first line treatment) and will not be referred to the RMHPs. Consequently, RMHPs only treat patient populations with a poorer prognosis. RCTs probably recruit MDD patients from both the GP population with a good prognosis and from the RMHP population (with a poorer prognosis) and in the "worst" case only from the first line population. The overall prognosis of RCT participants is therefore probably better than of RMHP patients. In order to optimize the generalizability of results from RCTs to "real life" (RMHP) psychiatric practice, it would be recommendable to conduct trials which include only patients who already went through GP or private practice treatment.

Therapist features

On a regular Friday, six a clock in the afternoon, Dr X. leaves his institution. It has been a busy week; at least twenty patients a day, staff meetings, resident supervision, two patients in severe crisis, an absent colleague who will probably be ill for a longer period, and a deadline for a report on a patient who had a complaint about his treatment. Dr X. cannot deny his feeling of tiredness and he starts to look forward to the moment that he will retire. Meanwhile, he feels a not-severe-but-nevertheless-nasty flu coming up. In his briefcase he has a brochure of a new and promising trial for MDD patients. On the cover of the brochure there is a smiling physician in a crispy white coat, who seems to be half the age of Dr. X.

Therapists who participate in trials might differ from daily practice clinicians in terms of workload, motivation, extent of updating training, and many other aspects. While for trial therapists the proper conductance of the treatment under investigation is their main goal, so to speak "real life" clinicians can be distracted by many other tasks than state of the art treatment of MDD patients. For instance, "real life" clinicians may have very limited time per patient as a result of a caseload that is too large. Furthermore they often have to stand in for absent colleagues, perform instant assessments of so called crisis patients, and sometimes also have managerial tasks. And all this in between their therapies for MDD patients…Furthermore, clinicians are probably very dedicated to their patients, yet perhaps headstrong when it comes to strictly following the protocol described in the treatment guidelines. Therapists who participate in trials might be more motivated for the treatment under investigation than "real life" clinicians. Maybe it is even so that especially highly skilled or specialized clinicians participate in trials. All these factors contribute to differences between RCT therapists and "real life" clinicians. Motivation, protocol adherence, extent of education, time per patient, and experienced workload are all factors likely to be associated

with treatment outcome. So, if *patient selection* is not the (only) answer to the difference between efficacy and effectiveness, *therapist selection* may well be one of them.

Differences between trial setting and daily practice setting

As a response to a shimmering trial brochure that calls for participants, Dr X. sends in Mr. Z. for participation. Mr. Z. is a 45 year old patient who suffers from MDD. He is a little bit sceptical about the results of antidepressant therapy, since his cousin and his neighbor did not improve on this medication. The trial therapist convinces Mr. Z. that the trial antidepressant is very new and promising. He explains the procedure of the trial to Mr. Z. and tells him that he will have a chance to either receive this new drug or a placebo. Mr. Z. is persuaded to participate. Without knowing (a double blind procedure) Mr. Z receives a placebo. The inspired trial therapist sees Mr. Z. every week during the follow-up time of the trial, which is 8-12 weeks. After 3 months, treatment results are assessed and the trial therapist says goodbye to Mr. Z. He thanks Mr. Z. for his willingness to participate and for his contribution to the development of treatment of MDD.

Every treatment has a placebo effect: the mere fact that the patient is receiving treatment has a beneficial effect. The aim of RCTs is to prove that the active drug under consideration has a significantly larger effect than a placebo (the between group efficacy, see above). In clinical practice the placebo effect also contributes to the overall treatment effect. It is likely, that the placebo effect in trials is larger than in daily practice. We will provide some arguments why this might be the case:

A proportion of participants in RCTs will spontaneously recover (like in daily practice) during participation. Spontaneous recovery will augment the proportion of patients who reach remission in a trial, while this effect cannot be attributed to the investigated treatment. In one meta analysis, spontaneous recovery was estimated to constitute one third of the placebo effect [16]. In daily practice, patients who recover spontaneously will probably not enter treatment or will drop out prematurely. They will not enter a ROM follow up assessment and therefore do not contribute to treatment outcome in ROM. Furthermore, participants in trials (and clinicians as well) have the feeling that they are treated in a special, new and promising way. This belief might contribute to improvement in RCTs and is called the Hawthorne effect [17,18]. Finally, as discussed above, in trials much effort is put in optimizing both patient's and clinician's protocol adherence. Protocol (or guideline) adherence seems to be positively associated with treatment success [19-25]. One specific aspect of protocol adherence is the frequency of follow up visits. In RCTs, frequency of appointments is closely monitored, while in daily practice appointments are sometimes cancelled by the clinician or patients for reasons of illness or otherwise. As a result of that, patients in RCTs have more regular and more frequent follow up visits. In a meta analysis on the therapeutic effect of follow up assessments in AETs, it was found that extra follow

up visits were associated with better treatment outcome and that the therapeutic effect of follow up assessments represents about 40% of the placebo response in AETs [26].

Study features

Therapist A conducted a trial with a new type of psychotherapy. She was very enthusiastic about this type of treatment, but unfortunately after a lot of effort it turns out that the results are disappointing. The effect of the new psychotherapy was comparable to that of treatment as usual. What are the chances that this therapist will lose her motivation and the results will end up in her top drawer? And if not, what are the chances that her negative results will be published in a prominent psychotherapy journal?

Negative findings are reported a lot less often than positive findings. This is the so called *publication bias* [27]. Nowadays all medication trials have to be made available in public registers that are available to everyone (e.g. Nederland's Trial Register) ahead of the start. However, treatment guidelines are based on articles published in scientific journals. Due to publication bias, efficacy may be overestimated. This may partially explain the difference between efficacy and effectiveness. Recently, several methodologies have been developed for meta-analyses in order to adjust to some extent for publication bias. It also has been suggested that the efficacy in MDD trials is exaggerated due to so called *rater bias*. The severity of MDD might be somewhat inflated by participating therapists at the beginning of the trial. At the same time, severity rating of the depression might be somewhat deflated at the end of the trial. If so, treatment success of trials (which can be the pre-post treatment difference) might be exaggerated and thus contribute to the difference between efficacy and effectiveness. However the extent of rater bias in MDD trials is still unknown. In one study rater bias was found to occur in MDD trials, yet its extent was too small to invalidate the results of the trials [28].

Minor depression, prior antidepressant use and having a paid job: implications of our findings

Although modest, we found that the exclusion of patients with mild depression, patients who used antidepressants prior to psychotherapy and patients without a paid job, improved treatment outcome in the remaining patient group. Although our most striking overall finding was a negative (absence of association) one: "In our "real life" patients, being eligible (meeting all criteria) for RCTs was *not* associated with a better treatment outcome", we didn't want to leave our positive findings undiscussed. The influence of exclusion of patients who do not meet the baseline severity threshold, who use antidepressants prior to their psychotherapy and who do not have a paid job is described in detail in the chapters 3, 4 and 6. Below we summarize our main findings.

We found that exclusion of ROM patients who suffered from minor depression (baseline severity of less than 18 on HAMD17) lead to a larger proportion of patients who reach remission (OR 2.0; 95% confidence interval 1.3–3.1). This association was found for psychotherapy, antidepressant treatment and a combination of both. As mentioned before, AETs often use a baseline severity threshold of HAMD17 ≤18 as exclusion criterion. We also found that exclusion of patients who have a baseline severity less than 14 on the HAMD17 (the threshold used in PETs) lead to more improvement in the remaining patients (β=7.23; 95% confidence interval 5.31–9.11).

We found that mild to moderate depression is very common in routine clinical practice (42% of the patients do not meet the AET severity threshold and 22% do not meet the PET severity threshold). Why this specific group rarely reach remission in their first treatment step is still unclear. Maybe these patients more often have a chronic mild depression instead of episodes of more severe MDD, and therefore have a different prognosis. It is also possible that these patients have other traits that differ from more severe MDD patients, such as lack of optimism as a personality trait. Future research is recommended on the characteristics of the large group of "real life" patients suffering from minor depression. To what extent the results of RCTs are generalizable to this group also needs to be further explored.

In addition, we found that exclusion of patients who used antidepressants prior to psychotherapy enlarges the extent of improvement of PETs (β=7.62; 95% confidence interval 1.94–13.30). These patients probably do not or only partially respond to medication, often prescribed by their GP or in private practice. As mentioned above in this chapter, these patients might have a worse treatment prognosis, than the ones who did not go through another treatment prior to their psychotherapy. Our finding accentuates that it would be recommendable to conduct trials which include only patients who already went through GP or private practice treatment, in order to optimize the generalizability of results from RCTs to "real life" (RMHP) psychiatric practice.

We compared demographic characteristics of the groups. We found a substantial difference in the proportion of patients employed at time of participation. 68% of the RCT participants had a paid job, while only 34% of the ROM patients were working at the time of treatment. ROM patients who were working had better treatment outcome than patients who were not, irrespective of the baseline severity of their depression (OR 1.76; 95% confidence interval 1.2–2.6 for the proportion of MDD patients who respond and OR 1.85; 95% confidence interval 1.2–2.8 for the proportion of patients who reach remission). In chapter 6 we showed that having financial security is probably not the aspect of having a job that contributes to treatment success. We recommend further research on which aspects of employment contribute to treatment outcome of MDD patients. The results of this future research can be used in the development of new MDD treatments or improvement of the existing ones by increasing the attention for the role of social factors in MDD treatment.

Limitations of our project

There were three major limitations to our research: Firstly, the lack of consistency in efficacy trials with respect to type of instruments, definition of outcome and use of exclusion criteria. Secondly, the missing data and large loss to follow up in the ROM data that is inherent to research in clinical practice. Lack of routinely collected treatment information and data on life history in the ROM data forced us to rely on data from extensive charts review. Thirdly, although our results seem to be representative for "real life" MDD patients, there are also limitations in the generalizability of our results. In the next paragraphs, we will first discuss the implications of the lack of consistency in RCTs, discuss the limitations of working with ROM data, and finally we will critically review the limits of the generalizability of our results.

Research on research: limitations in estimating efficacy

We investigated the efficacy of antidepressant treatment, individual psychotherapy and combination treatment. We studied the estimation of efficacy in RCTs and found an inconsistency in the use of instruments to assess depression severity. We also found an inconsistency in the definition of outcome: response is consistently defined as a 50% reduction of symptoms, but remission is defined by different cut off scores. Furthermore, we found that PETs are inconsistent in their use of exclusion criteria. These inconsistencies in the underlying data might compromise the validity of the aggregated efficacy estimates that are given in meta-analyses.

In addition, AETs and PETs have a different manner in evaluating treatment outcome, due to a different research tradition. The difference of defining outcome between AETs and PETs did not hinder our analysis, but it somewhat diminished the comprehensiveness of our results for clinicians, since we had to compute outcome in line with AETs as well as PETs. In table 1, we provide an overview of the instruments and definitions of outcome used in the meta- analyses included in our study. In the frame we describe the inconsistencies in instruments and outcome definition and their implications for our results in detail. The comparability of efficacy estimates, in meta- analyses but also in the comparison with "real life" cohorts would benefit greatly from more consensus on the instruments and the eligibility criteria for AETs and PETs. Finally, within our selection of AETs and PETs, for the exploration of eligibility criteria in PETs (chapter 4) and the reporting of sociodemographic/ socioeconomic features in AETs as well as PETs (chapter 5), many more AETs were available than PETs. For AETs we therefore limited our search to high impact journals, while we included all PETs within the same time frame. Although our selected AETs and PETs were similar with respect to countries of origin and timeframe, there is a slight possibility that our methodology of RCT selection has introduced some selection bias.

Table 1. Instruments and definitions of outcome in meta analyses.

		Type of meta-analysis	RCTs that used the following instruments were included	Definition of outcome: Response	Definition of outcome: Remission	Definition of outcome: Effectsize
1	Kasper 1997	AETs	HAMD (17 item version)	50% reduction	HAMD ≤7	-
2	Bech 2000	AETs	HAMD (17 item version)	50% reduction	-	-
3	Storosum 2001	AETs	HAMD (both 17 and 21 item version)	50% reduction	-	-
4	Steffens 1997	AETs	HAMD (version not specified)	50% reduction	-	-
5	Montgomery 2001	AETs	HAMD (17 item version)	50% reduction	HAMD ≤8	-
6	Beasley 2000	AETs	HAMD (17 item version)	50% reduction	HAMD ≤7	-
7	Einarson 1999	AETs	HAMD (version not specified) MADRS	50% reduction	-	-
8	Stahl 2002	AETs	HAMD (21 item version) MADRS	50% reduction	-	-
9	Nelson 1999	AETs	HAMD (version not specified) MADRS	50% reduction	-	-
10	Thase 2001	AETs	HAMD (both 17 and 21 item version)/ MADRS	50% reduction	HAMD17 ≤7; HAMD21≤7/≤8/10; HAMD17≤10 +CGI=1, MADRS<10	-
11	Thase 1997	PETs COMs	HAMD (17 item version)	-	HAMD<7	-
12	De Maat 2007	PETs	HAMD BDI	-	HAMD <6/<7/<8; BDI <9/<10	-
13	Minami, 2007	PETs	BDI	-	-	Δ Mean BDI pre-posttreatment / SD pre-treatment
14	Thase, 2005	AETs	HAMD (21 item version)	-	≤7 on first 17 items on HAMD21	-
15	Wexler, 1992	COMs	BDI	-	BDI≤16/23; Raskin[2] ≤9	

AET: antidepressant efficacy trial. PET: psychotherapy efficacy trial. COM: combination treatment trial; antidepressants + individual psychotherapy. RCT: randomized controlled trial. HAMD: Hamilton Depression Rating Scale. MADRS: Montgomery Asberg Depression Rating Scale. BDI: Beck Depression Inventory. CGI: Clinical Global Impression scale Raskin: Raskin Depression Scale

Cicchetti DV, Prusoff BA: Reliability of depression and associated clinical symptoms. Arch Gen Psychiatry 1983, 40: 987-990.

Δ: difference pre-post treatment. SD: standard deviation. - : Definition of outcome is not used in meta-analysis.

Inconsistencies in the use of instruments in RCTs

In AETs, the most commonly used severity scale is the HAMD [9], especially in trials from the United States. The MADRS [29] is also often used in AETs, especially in European trials, sometimes as primary outcome measurement, often as secondary instrument. In our selection of meta analyses on AETs, they all used the HAMD. Yet, two different versions of the HAMD (17 and 21 items) are used. Both versions are validated, but how a cut off score for remission on one version relates to a cut off score on the other version is not clear. In ROM the MADRS is used. The fact that we had to convert the MADRS to HAMD scores in all our analyses, might have influenced our results on the efficacy-effectiveness difference. In order to give the most reliable estimate of HAMD scores, we used three equations [30,35,36] to convert our MADRS scores. We found that two equations yielded the similar results [30,36] and we performed a validity check with another method for conversion: the Item Response Theory [32]which also yielded similar results. We therefore expect little limitations to our analyses due to the fact that HAMD is not used in ROM.

Inconsistencies in the use of cut offs for remission in RCTs

All meta analyses on AETs used the same definition of response and therefore we did not encounter difficulties in the efficacy–effectiveness comparison. However, most patients in ROM suffered from mild to moderate depression, which lead to very similar proportions of response and remission. Therefore, we did not report separately on the efficacy-effectiveness difference for response (chapter 2). The definition of remission varies between AETs and some meta analyses include trials with different definitions of remission. In our selection of meta analyses four different cut off scores to define remission were used. For the computation of the effectiveness of MDD treatment in "real life" we used a stringent (and scientifically investigated) cut off of MADRS ≤10 [30], which equals a score of 6.4 on the HAMD17 [31,32]. By using this stringent cut off score, it might be that we were too harsh in estimating the efficacy-effectiveness difference. In reality the efficacy-effectiveness difference might be a little smaller, especially for the efficacy-effectiveness difference in meta analyses that included only trials that used a less stringent cut off [33,34]. However, the most often used cut off score for remission (HAMD17 score of 7) in the meta analyses is, to our opinion, close enough to our definition of remission in "real life" (MADRS ≤10 ≈ HAMD17≤6.4)) to give a reliable estimate of the difference between efficacy and effectiveness.

Inconsistencies in the use of exclusion criteria in PETs

We found that PETs are not consistent in the exclusion criteria they use. Only 4 of the 38 criteria were used in 75% of the papers (chapter 4). This, of course, hampers the comparability of PETs and thus the reliability of meta analyses of PETs (and the comparability of PETs with AETs). It also has consequences for the interpretation of the results described in chapter 4. Firstly, calculating the overall efficacy of PETs as is done in three of our selected meta analyses [37-39] while the comparability of PETs is low, raises questions about the reliability of the results from these meta analyses. Therefore, the reliability of our results on the efficacy-effectiveness difference might likewise be jeopardized. Secondly, it was impossible to take all exclusion criteria into account when we investigated which "real life" patients would have been eligible for PETs. We restricted ourselves to the four most consistently used criteria, making the comparison of treatment effects in eligible and non-eligible patients just an approximation.

Research on ROM data: limitations in estimating effectiveness

The ROM data were gathered in clinical practice, as part of the routine diagnostic and treatment processes. Although such data have the advantage of offering insight into the vicissitudes of "real life" patients, they also have limitations just because of these vicissitudes. First of all, data integrity is not guaranteed. By using computers with touch screens and software that makes it impossible to skip a question in a questionnaire and by having test nurses supervising the filling out of the questionnaires, we tried to make the data as complete as possible. However, it was clinical practice, not a research project in which double checking of data and data gathering are the standard procedure. Thus incompleteness was inevitable. Also the large number of questionnaires may have impeded completeness. We addressed the problem of missing data as good as possible by using elaborate statistical methods (MICE, multivariate imputation by chained equations, [40]). Second, in the period in which the data for our project were gathered, the follow up assessments in ROM were not organized properly. The consequence is an almost 50% loss to follow-up. In the relevant chapters of this thesis we discussed how we tried to handle this loss. On the other hand, a large loss to follow up may be inherent to studies with a naturalistic design: STAR*D had reached a loss-to-follow-up of 48% in step II of the study. Third, in ROM data on the history of the patient's life and his illnesses are rudimentary. Unfortunately, as those data are also not available in a useful digital format, we had to depend on an extensive chart review. All these factors will have reduced the reliability of the data. However, they are more extensive and relate to a larger number of patients than in any other project. Therefore, we felt that our data are a significant contribution to this new field of research.

Generalizability of our results: limitations

In this thesis, we explored the outcome of antidepressant and psychotherapeutic treatment of MDD from baseline assessment to the first follow up assessment. We have not addressed patient selection and its influence on outcome of RCTs on combination treatment, as combination treatment is a second step in the treatment algorithm of MDD.

As mentioned earlier, all ROM patients at the RMHP are referred by either their GP (most often) or by a psychiatrist working in private practice. Many patients already underwent treatment for their MDD prior to referral. Our results are therefore only generalizable to outpatient clinics that treat similar patients. The generalizability of our results to private practices, GP practices and mental health providers who treat only or merely patients that are treatment naïve (and who probably have a more favorable prognosis) or patients who are non responders to several therapies (so called third line institutions) is most likely limited.

The meta-analyses that we used in this project were carried out in the United States of America and Europe. These studies included a predominantly white patient population. Our ROM population also is a predominantly white patient population treated in Western psychiatric practice. We do not know whether our results are also valid in other cultures. Neither do we know to what extent they apply to non-Western immigrants in the Netherlands who were unable to fill out the questionnaires.

Future directions in effectiveness research and opportunities for clinical innovation

In this last section of the discussion we will present some recommendations for future research in line with our project and also for clinical development.

We will start with recommendations for future research.

- As described earlier, our loss to follow-up was considerable. From personal communications with other centers using ROM it is clear that this is a nearly universal problem. The large loss to follow up in the STAR*D trial also emphasizes the problem of loss to follow up in research done in clinical practice. Of course, the loss to follow up could be decreased by a better organization. Probably the covenant with the insurance companies to increase the proportion of patients with follow up data may help. However, the high loss to follow up should also become a focus of research. Almost one third of these lost to follow up patients remain in treatment, so do these patients refuse to participate in ROM or do clinicians forget to sign up their patients? Future research will have to focus on reasons why patients do not participate in follow up assessments. And the other two thirds of the patients? Did they recover and then disappear? Or were they unsatisfied with their treatment and no longer showed up? It is remarkable that the urge to investigate these topics is not felt widely. Perhaps patients who are lost to follow up have specific features such as a common social background, more co morbid disorders or specific personality traits. From our lost to follow up analysis, we learned

that although the patients who were lost to follow up were very similar to the ones who were not, especially single male MDD patients suffering from co morbid post traumatic stress disorder were at risk of dropping out and being lost to follow up. Future research might reveal specific subgroups that are at risk for drop out or loss to follow up and need a specific approach to stay in treatment and have a proper evaluation of it. Also more research is needed on the side of the therapists: are there specific professional groups that do not support ROM? And what do they need to feel the need for routinely systematic evaluation of their treatment?

- Research on treatment effectiveness and benchmarking requires large databases. Therefore, it is important that the ROM of mental health care centers use, as far as possible, the same questionnaires and procedures. In the current financial crisis, many policy makers need/tempt to make stringent cutbacks in the budgets of mental health care. One way to reduce the costs of ROM is to reduce the number of instruments in ROM as much as possible. This, however, may seriously jeopardize the usefulness of ROM data as a reliable instrument for the evaluation of treatment progress in clinical practice. Furthermore, it certainly jeopardizes the usefulness of ROM data for scientific research. A discussion about what the necessary ingredients of ROM are, is necessary. The data of the Leiden Routine Outcome Monitoring Study may be helpful to provide this discussion with data, i.e. by the exploration of the validity of key items in the available instruments and the possibilities of answer-steered exposure to new items of questionnaires (patients do not fill out complete questionnaires, but will get new items based on their response to the former ones).
- Further research on the influence of factors in which AETs and PETs on one hand and "real life" patient cohorts on the other hand differ, should be continued. More specifically, data not included in this study, for instance on earlier treatments and patient history, should be included. Also, then, replication studies on our findings can be carried out, preferably in real life cohorts with more complete data and less loss to follow-up.
- We investigated MDD. It would be useful to extend this type of research to other disorders, for instance anxiety disorders. Such research would elaborate which problems are unique for MDD and which are general.

Is our finding of modest effects of the first step in evidence based MDD treatments a reason to discard the guidelines, throw away evidence based medicine and go back to experience based medicine? Back to the "good old days" where individual doctors knew best for individual patients and where clinicians acted on personal experience? No. Research indicates that there is a positive association between the introduction of evidence based therapies in daily practice and the improvement of MDD treatment, yet its relation is still not unmistakably clear. It is time to answer the question that was asked by A.J Rush in 1993: "Clinical Practice Guidelines: good news, bad news, or no news?"[41]. Many researchers have

tried to answer this significant question, and it is a difficult one to answer. Many factors are involved in treatment in daily practice and therefore daily practice is very complicated to address scientifically [19]. ROM is a very promising and valuable methodology to get insight in the many aspects of routine psychiatric practice. So, indeed, effectiveness in daily practice is not as positive as we hoped for. But go back to experience based treatment? No one knows what treatment results were before the introduction of evidence based medicine. And no one will ever find out, because experience based treatments cannot be explored in terms of effectiveness, since they differ between each patients and nothing is recorded automatically. So one of the big yet bitter advantages of the introduction of guidelines is that we now *know* that effectiveness is currently not as good as the promising results from RCTs. ROM may provide data to improve in an evidence based way the treatment results in clinical practice, e.g. by the future possibilities to identify patients who are at risk of non response or to define subgroups of patients that respond better to a certain type of treatment.

Are the modest MDD treatment results in daily practice a reason for panic or despair, then? A reason to become depressed? A reason to cut down the budget on mental health? We don't think so. The age of evidence based medicine went hand in hand with the age of optimistic belief in antidepressant treatment, efficacy trials on ssri's and an enormous increase in the prescription of antidepressants. Among others, pharmaceutical industries conducted efficacy trials on antidepressants and showed that depression is a treatable disorder. Those days of optimism are over. Antidepressants seem not be as effective as was believed [42], not even in the short term, 6-8 weeks follow-up trials in which no effort is spared to optimize adherence, and in which only patients with moderate to severe MDD are treated. From our results and from those of the STAR*D trial [10,11], it is clear that in daily practice even short term treatment of MDD is hard, and the results modest. In addition, it has become evident that depression is a chronic illness [43-45], which remits and recurs, and rarely disappears.

Depression causes a lot of suffering, some patients who suffered both from very severe "somatic" illnesses or terrible personal losses and major depressive disorder, stated that their depression was the worst. The loss of hope, a continuous feeling of worthlessness and/or despair, the inability to participate in daily life in the broadest sense of the word together with all the physical complaints that may occur when one suffers from major depressive disorder, surely makes MDD a disease that justifies all efforts from patients, caregivers, clinicians, researchers, and mental health policy makers. MDD is an expensive disease with respect to direct costs on the health care budget (not only the mental health care budget) and indirect costs with respect to absenteeism. Depression is the leading cause of disability and the fourth leading contributor to the global burden of disease according to the World Health Organization (https://www.who.int/en). MDD is a very serious medical issue, like other chronic diseases such as diabetes or chronic obstructive pulmonary disease (COPD), not a temporarily suffering from worries that will go away after a good talk with

your neighbour. It is of importance that the people in charge of the mental health care budgets, the policymakers, the government and the minister of Public Health take notice of the complexity, severity and chronicity of MDD.

Having learned that we, as clinicians and researchers, have to be modest about the prognosis of patients suffering from MDD who seek treatment for it, we cannot just sit and wait... We can improve treatment adherence of patients and clinicians and we can develop staging and profiling of MDD. This discussion will be closed by elaborating on these three topics.

Firstly, we can ameliorate our methods to improve treatment adherence of patients. Many new developments may improve adherence: e-health, apps with medication instruction, sms alerts for medication, technical devices that help patients to monitor changes in their mood by providing feedback several times a day, and collaborative care (an integrated approach of the biological, social and psychological aspects of MDD). For better treatment adherence, we have to invest in the education of patients suffering from MDD. It has been proven that informing patients about the nature of their disease and its treatment, the duration, the expected results and time span, the expected investment of the patients and possible side effects of the treatment, which accounts for both pharmacotherapy and psychotherapy will improve adherence [46,47]. It is hard to tell a patient who just got out of a period of feeling worthless and guilty, who had nights without sleep, days without energy or appetite and who nearly came to the edge of committing suicide, that it is likely that this illness will return, sooner or later. Nevertheless, education is a very important part of MDD treatment. Future research on the effect of improvement of treatment adherence on outcome in daily practice is highly recommended. Secondly, improvement of the protocol adherence of clinicians in daily practice might also lead to an increase of effectiveness. Due to a variety of reasons, clinicians in daily practice sometimes find it difficult to strictly follow the protocol (especially when it comes to the frequency of follow up contacts or taking blood levels of antidepressants). Further research on the association between protocol adherence and outcome is recommended [19]. In this project, we presented ROM as a valuable methodology to do scientific research in daily practice. Other potential benefits of ROM remained underexposed in this thesis so far, but need to be mentioned. In Rivierduinen, ROM was primarily designed to ameliorate the evaluation of the treatment of individual patients and patient groups. A systematic evaluation of treatment progress after each treatment step helps clinicians and patients to see whether they are on the right track, and what further steps need to be taken. If ROM is fully incorporated in daily routine, it can be a helpful tool for clinicians to remain adherent to their treatment protocol and to switch in time to a next treatment step in the protocols for MDD treatment.

Finally, depression treatment itself can be improved by so called staging and profiling (a specific therapy for a specific stage or subtype of the disease). At this moment, almost all patients suffering from MDD are treated the same way, either with antidepressants or with

psychotherapy (or a combination of both). The choice for either one of treatment modalities is based only on severity of the depression and the preferences of the patients. Yet, there are many clues that not all MDDs are the same. Within the disorder, different symptoms or symptom dimensions may have different etiology. Different symptom dimensions [48] have been demonstrated to be associated with different genetic pathways [Van Veen, in press], with differences in the dysregulation of the HPA-axis [49,50], and different types of childhood trauma [51] [Van Veen, submitted] and life events [Wardenaar, submitted]. Currently more and more results become available indicating that different subtypes of depression need different treatment. For instance in the STAR*D trial was found that specific genotypes together with co morbid anxiety disorders (in our ROM sample 43% suffered from co morbid anxiety and/or somatoform disorders) are associated with non-response to antidepressants [52]. Similar results were found in the Genome Based Therapeutic Drugs for Depression (GENDEP) study [53].Therefore, patients suffering from MDD should not all be treated in the same way, but with treatment tailor made for their type of depression. Future research should focus on those tailor made treatments.

REFERENCE LIST

1. 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.
2. Zetin M, Hoepner CT: Relevance of exclusion criteria in antidepressant clinical trials: a replication study. *J Clin Psychopharmacol* 2007, 27: 295-301.
3. Partonen T, Sihvo S, Lonnqvist JK: Patients excluded from an antidepressant efficacy trial. *J Clin Psychiatry* 1996, 57: 572-575.
4. Rothwell PM: External validity of randomized controlled trials: "to whom do the results of this trial apply?". *Lancet* 2005, 365: 82-93.
5. 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.
6. 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.
7. 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.
8. 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.
9. Hamilton M: Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967, 6: 278-296.
10. Thase ME, Friedman ES, Biggs MM, Wisniewski SR, Trivedi MH, Luther JF *et al*.: Cognitive therapy versus medication in augmentation and switch strategies as second-step treatments: a STAR*D report. *Am J Psychiatry* 2007, 164: 739-752.
11. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L *et al*.: Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 2006, 163: 28-40.
12. McPherson K: Do patients' preferences matter? *BMJ* 2008, 337:a2034. doi: 10.1136/bmj.a2034.: a2034.
13. Churchill R, Khaira M, Gretton V, Chilvers C, Dewey M, Duggan C *et al*.: Treating depression in general practice: factors affecting patients' treatment preferences. *Br J Gen Pract* 2000, 50: 905-906.
14. Dwight-Johnson M, Sherbourne CD, Liao D, Wells KB: Treatment preferences among depressed primary care patients. *J Gen Intern Med* 2000, 15: 527-534.
15. Persaud N, Mamdani MM: External validity: the neglected dimension in evidence ranking. *J Eval Clin Pract* 2006, 12: 450-453.
16. Kirsch I, Sapirstein G: Listening to Prozac, but hearing Placebo: a meta-analysis of antidepressant medications. In *How Expectancies shape Experience*. American Psychological Association; 1999:302-320.
17. Leonard KL: Is patient satisfaction sensitive to changes in the quality of care? An exploitation of the Hawthorne effect. *J Health Econ* 2008, 27: 444-459.
18. Kirsch I (Eds):Response Expectancy of Experience and Behavior. In *American Psychologist* 1985, 40: 1189-1202.
19. Fenema EMv, Wee Nvd, Bauer M, Witte C.J., Zitman FG: Assessing adherence to guidelines for common mental disorders in routine clinical practice. *International Journal for Quality in Health Care* 2011, 1-8.
20. Katon W, Von KM, Lin E, Walker E, Simon GE, Bush T *et al*.: Collaborative management to achieve treatment guidelines. Impact on depression in primary care. *JAMA* 1995, 273: 1026-1031.

21. Badamgarav E, Weingarten SR, Henning JM, Knight K, Hasselblad V, Gano A, Jr. *et al*.: Effectiveness of disease management programs in depression: a systematic review. *Am J Psychiatry* 2003, 160: 2080-2090.
22. Andrews G, Issakidis C, Sanderson K, Corry J, Lapsley H: Utilising survey data to inform public policy: comparison of the cost-effectiveness of treatment of ten mental disorders. *Br J Psychiatry* 2004, 184: 526-533.
23. Pirraglia PA, Rosen AB, Hermann RC, Olchanski NV, Neumann P: Cost-utility analysis studies of depression management: a systematic review. *Am J Psychiatry* 2004, 161: 2155-2162.
24. Neumeyer-Gromen A, Lampert T, Stark K, Kallischnigg G: Disease management programs for depression: a systematic review and meta-analysis of randomized controlled trials. *Med Care* 2004, 42: 1211-1221.
25. Trivedi MH, Claassen CA, Grannemann BD, Kashner TM, Carmody TJ, Daly E *et al*.: Assessing physicians' use of treatment algorithms: Project IMPACTS study design and rationale. *Contemp Clin Trials* 2007, 28: 192-212.
26. Posternak MA, Zimmerman M: Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials: meta-analysis. *Br J Psychiatry* 2007, 190: 287-292.
27. Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G: Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *Br J Psychiatry* 2010, 196: 173-178.
28. Petkova E, Quitkin FM, McGrath PJ, Stewart JW, Klein DF: A method to quantify rater bias in antidepressant trials. *Neuropsychopharmacology* 2000, 22: 559-565.
29. Asberg M, Montgomery SA, Perris C, Schalling D, Sedvall G: A comprehensive psychopathological rating scale. *Acta Psychiatr Scand Suppl* 1978, 5-27.
30. 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.
31. Zimmerman M, Posternak MA, Chelminski I: Implications of using different cut-offs on symptom severity scales to define remission from depression. *Int Clin Psychopharmacol* 2004, 19: 215-220.
32. 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.
33. Montgomery SA: A meta-analysis of the efficacy and tolerability of paroxetine versus tricyclic antidepressants in the treatment of major depression. *Int Clin Psychopharmacol* 2001, 16: 169-178.
34. Thase ME, Entsuah AR, Rudolph RL: Remission rates during treatment with venlafaxine or selective serotonin reuptake inhibitors. *Br J Psychiatry* 2001, 178:234-41.
35. Mittmann N, Mitter S, Borden EK, Herrmann N, Naranjo CA, Shear NH: Montgomery-Asberg severity gradations. *Am J Psychiatry* 1997, 154: 1320-1321.
36. Hawley CJ: Depression rating scales can be related to each other by simple equations. 1998.
37. de Maat SM, Dekker J, Schoevers RA, de Jonghe F: Relative efficacy of psychotherapy and combined therapy in the treatment of depression: a meta-analysis. *Eur Psychiatry* 2007, 22: 1-8.
38. Minami T, Wampold BE, Serlin RC, Kircher JC, Brown GS: Benchmarks for psychotherapy efficacy in adult major depression. *J Consult Clin Psychol* 2007, 75: 232-243.
39. Thase ME, Greenhouse JB, Frank E, Reynolds CF, III, Pilkonis PA, Hurley K *et al*.: Treatment of major depression with psychotherapy or psychotherapy-pharmacotherapy combinations. *Arch Gen Psychiatry* 1997, 54: 1009-1015.
40. Royston P: Multiple imputation of missing values: update. *Stata Journal* 2005, 5: 188-201.
41. Rush AJ: Clinical practice guidelines. Good news, bad news, or no news? *Arch Gen Psychiatry* 1993, 50: 483-490.

42. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R *et al*.: Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009, 373: 746-758.
43. Andrews G: Should depression be managed as a chronic disease? *BMJ* 2001, 322: 419-421.
44. Gask L: Is depression a chronic illness? For the motion. *Chronic Illn* 2005, 1: 101-106.
45. Judd LL, Akiskal HS, Maser JD, Zeller PJ, Endicott J, Coryell W *et al*.: A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders. *Arch Gen Psychiatry* 1998, 55: 694-700.
46. Trivedi MH, Lin EH, Katon WJ: Consensus recommendations for improving adherence, self-management, and outcomes in patients with depression. *CNS Spectr* 2007, 12: 1-27.
47. Vergouwen AC, Bakker A, Katon WJ, Verheij TJ, Koerselman F: Improving adherence to antidepressants: a systematic review of interventions. *J Clin Psychiatry* 2003, 64: 1415-1420.
48. den Hollander-Gijsman ME, de BE, van der Wee NJ, van Rood YR, Zitman FG: Distinguishing between depression and anxiety: a proposal for an extension of the tripartite model. *Eur Psychiatry* 2010, 25: 197-205.
49. Wardenaar KJ, Vreeburg SA, van VT, Giltay EJ, Veen G, Penninx BW *et al*.: Dimensions of depression and anxiety and the hypothalamo-pituitary-adrenal axis. *Biol Psychiatry* 2011, 69: 366-373.
50. Veen G, van Vliet IM, DeRijk RH, Giltay EJ, van PJ, Zitman FG: Basal cortisol levels in relation to dimensions and DSM-IV categories of depression and anxiety. *Psychiatry Res* 2011, 185: 121-128.
51. van Reedt Dortland AK, Giltay EJ, van VT, Zitman FG, Penninx BW: Personality traits and childhood trauma as correlates of metabolic risk factors: the Netherlands Study of Depression and Anxiety (NESDA). *Prog Neuropsychopharmacol Biol Psychiatry* 2012, 36: 85-91.
52. Ising M, Lucae S, Binder EB, Bettecken T, Uhr M, Ripke S *et al*.: A genomewide association study points to multiple loci that predict antidepressant drug treatment outcome in depression. *Arch Gen Psychiatry* 2009, 66: 966-975.
53. Uher R, Perroud N, Ng MY, Hauser J, Henigsberg N, Maier W *et al*.: Genome-wide pharmacogenetics of antidepressant response in the GENDEP project. *Am J Psychiatry* 2010, 167: 555-564.