To fail or not to fail : clinical trials in depression

Sante, G.W.E.

Chapter

5

The missing link between clinical endpoint and pharmacological receptor systems in depression

Gijs Santen, Meindert Danhof, Oscar Della Pasqua

Submitted for publication

ABSTRACT

The basic clinical trial paradigm for the assessment of antidepressant efficacy has not changed in the past 50 years. Despite evidence of the relevance of different aspects of the disease and increased understanding of the complex neurochemical processes associated with mood disorders, global disease severity measures such as the Hamilton depression rating scale (HAMD) remain the gold standard in clinical depression trials. In the light of the development of antidepressants with new mechanisms of action, it is of interest to investigate the behaviour of the HAMD to different mechanisms of action (MoA). In this paper we propose the use of novel graphical methods to investigate the presence of bias of the HAMD to specific mechanisms of action.

A total of 5035 patients from 11 clinical studies in which placebo, TCAs, SSRIs and anticonvulsant drugs were administered to patients with major depressive disorder have been retrieved from GSK's clinical trial database. Based on a dichotomisation of patients into responders or non-responders, two types of graphical representations were used to describe (1) the rate of response for each individual item, yielding score-distribution over time separately for responders and non-responders, and for each mechanism of action, and (2) the extent of response by evaluating the contribution of each item to the total change in HAMD at the last observation.

Our findings reveal that the individual items of the HAMD scale are insensitive to differences in mechanism of action. The time course of response differs only between responders and non-responders in the population. Furthermore, there is no difference in the contribution of individual items to the total change in HAMD at completion of treatment for the different classes of drugs. Interestingly, variability in the contribution of individual items is considerably larger in non-responders than in responders.

This work provides evidence that the HAMD is not an appropriate clinical measure for differentiating compounds with distinct mechanisms of action. We recommend using the proposed graphical analysis to detect if a new MoA may affect individual items of the HAMD specifically, rather than relying on the total HAMD change. However, a mechanism-based approach is required that enables assessment of multidimensionality of symptoms and signs. Composite endpoints that reflect underlying mechanisms of action need to be developed and validated without more ado.

INTRODUCTION

Regardless of the high failure rate of clinical trials in depression (Khan *et al.*, 2002b), the concepts underlying clinical trial design have not changed in the past 50 years. Only summary measures of improvement and disease severity continue to be used as primary endpoints in the evaluation of antidepressants, despite evidence of the relevance of differential affective and behavioural components of the disease and increased understanding of the complex neurochemical processes associated with mood disorders (Juckel *et al.*,

2007; de Kloet *et al.*, 2007). Examples of these measures are the Hamilton depression rating scale (HAMD) (Hamilton, 1960) and the Montgomery Asberg depression rating scale (MADRS) (Montgomery and Asberg, 1979). Inevitably, the use of such scales regards antidepressants as drugs that treat 'depression' as a unidimensional, unitary disorder. This notion contrasts with current research efforts, which focus on the specificity of action at selected receptor systems. It is conceivable that novel, antidepressant drugs may affect only specific components of this heterogeneous disease, and global disease assessment scales may therefore fail to detect these effects. A componential model for the assessment of depression has been recommended which takes into account the different aspects of the symptomatology, such as mood, behaviour, cognitive and somatic components (Katz, 1998).

Within R&D, pharmaceutical industry endeavours to differentiate compounds that provide better efficacy and safety profiles. In the past decade, anticonvulsant drugs such as lamotrigine have been tested for their efficacy in unipolar and bipolar depression (Green, 2003). Recently, NK1-antagonists have been shown to be viable drugs in depression (Kramer *et al.*, 2004). Undoubtedly, other drugs with new mechanisms of action will follow (Moret, 2003; Pacher and Kecskemeti, 2004). The success of such attempts depends upon the sensitivity of available clinical endpoints. Various meta-analyses have failed to show differences in efficacy between classes of antidepressants using global disease severity scales (Papakostas and Fava, 2006, 2007a,b) such as the HAMD or MADRS. In contrast, differences between antidepressants have been found with regard to their adverse event profile, such as reported by Kennedy *et al.* (2000).

In other therapeutic areas developments have occurred to identify how differences in pharmacological properties of a drug may be correlated to changes in clinical measures. In particular, it is essential to establish whether the relationship between mechanism of action and clinical response is univocal. Some examples include the link of the GABAergic receptor complex to EEG waves (Mandema and Danhof, 1992; Visser *et al.*, 2003), the direct correlation between clinical extra-pyramidal symptoms in Parkinson's disease and the dopaminergic receptor system (Volkow *et al.*, 1998), the link between muscarinic receptor blockade and mucus hyper-secretion in COPD (Gosens *et al.*, 2006) and the relation between D2 dopamine receptor activation and positive psychotic symptoms in schizophrenia (Pani *et al.*, 2007).

Sadly, the aforementioned advancement has not occurred in depression. Most imaging studies focus on biomarker properties of PET technology, rather than dissecting the correlation between differences in receptor occupancy and its potential correlation with individual items and total score of clinical rating scales. To our knowledge reports on this field of psychiatric research so far remain qualitative in nature.

Whilst the Hamilton depression rating scale (HAMD) has been criticised extensively (Bagby *et al.*, 2004; Bech and Rafaelsen, 1980; Bech, 2006), it remains being used as a global disease severity measure and is the primary endpoint in most clinical trials, However, one must consider that in 1960, when the HAMD was first published, only tricyclic

antidepressants (TCAs) were available for treatment. The first new mechanism of action (MoA) that became available for the treatment of depression was specific serotonin reuptake inhibition (SSRIs) in the 80s. Even though the HAMD was not devised to monitor change upon treatment, but rather as a diagnostic tool, it has been suggested that the HAMD is more sensitive to detect TCA effect compared to SSRI effects. Some papers have tried to investigate this so-called bias, but their results are contradictory (Khan *et al.*, 2004; Moller, 2001; Nelson *et al.*, 2005a). It is important to elucidate any such effect, since it may hinder drug development in depression. For example, a drug with sedative effects may change the insomnia-related items of the HAMD and therefore lead to a significant treatment effect in a depression trial. If the item *depressed mood* is not changed by this drug, one may question whether this drug should be classified as an anti-depressant. Inversely, a drug which performs well on the item *depressed mood* but has no effect on the other items may have anti-depressant effects but may not result in a significant effect when the HAMD is used as clinical endpoint.

An intrinsic difficulty is encountered when trying to determine if the HAMD favours one MoA over the other: it may well be that the favoured MoA simply is a better anti-depressant! Since there is no external validation in the form of an independent benchmark, one has to be careful not to end up in circular arguments. Fortunately, in this case the problem is part of the solution. We can use the original intent of the HAMD, i.e., assessing the severity of depression in a given patient. This should not depend on the particular anti-depressant taken. In a previous investigation we have proposed a graphical method to explore the sensitivity of individual items of HAMD and MADRS using the difference between responders and non-responders instead of the traditional comparison between active treatment and placebo (chapters 3 and 4). This approach resulted in a new response-based subscale (HAM-D_7) (chapter 3), consisting of the *suicide* item and the items previously included in the Bech and Rafaelsen (1980) HAM-D_6 (*depressed mood*, *feelings of guilt*, *psychic anxiety*, *work and interests*, *somatic symptoms general* and *retardation*). A comparison between the full HAMD, Bech HAM-D_6, the response-based subscale and the MADRS showed that the HAMD subscales were more sensitive to drug effect than the MADRS (chapter 4).

It is plausible to assume that HAMD items previously identified as sensitive to response remain so irrespective of the MoA. Likewise one could expect insensitive items not to be affected by differences in pharmacological properties. This hypothesis raises the question whether novel drugs with distinct MoAs and specific modulatory effect on a sensitive or insensitive item will ever be differentiated in the current efficacy trial paradigm. In this paper we will present novel graphical approaches to evaluate whether the HAMD is sensitive to differences in MoA. Consequently, a bias, both positive and negative, to the drug under investigation may be revealed. We also anticipate that clinical trial design for these new mechanisms of action may benefit from the methods proposed here.

METHODS

Study data

A total of 5056 patients from 11 placebo-controlled, randomised clinical trials in major depressive disorder were retrieved from GlaxoSmithKline's clinical database. Inclusion and exclusion criteria were similar between studies, and for all studies patients were required to be diagnosed with major depressive disorder and to abstain from any other concomitant antidepressant medication during the trial. Further information on the studies and references to publications are presented in table 1. All information can also be retrieved from the GSK clinical trial register (http://ctr.gsk.co.uk). All studies were performed in adults.

In addition to placebo, data on three different mechanisms of actions were selected for the purposes of our analysis. Imipramine and desipramine were the representatives of the tricyclic antidepressants (TCAs), fluoxetine and several formulations of paroxetine represented the serotonin-specific re-uptake inhibitors (SSRIs) and lamotrigine was the sole compound in the anticonvulsant (AC) class. To account for the possible confounder of differences in systemic exposure, data from all doses (i.e., therapeutic and sub-therapeutic dose levels) were included in the analysis. Since treatment duration and visit frequency were different across studies, we have chosen to normalise the denominator for assessment times by grouping HAMD scores in weeks 7 and 8 with those in week 9. We have also excluded all observations in week 5.

Sensitivity of the HAMD to mechanisms of action

In order to assess any differential effects of MoA on the HAMD, two approaches were used. In the first approach, the study population was split in a responder and non-responder subset. Full details of the method have been published previously (chapter 3). Briefly, patients were considered responders if their HAM-D_{17} was reduced at least 50% from the baseline value at any time during the trial. All observations were grouped by week of visit and the time course of response was then analysed by showing the proportion of patients scored with each possible value for the individual item (Jonsson, 2004). This procedure enabled us to visualise the time course of each item for different MoAs, separately for responders and non-responders.

In addition to the indication about the response rate during the course of treatment, which can be derived from the aforementioned temporal patterns, the second approach proposed in this manuscript provides evidence for the extent of response at completion of treatment. For that purpose, only the first and last observed HAMD score were used for each patient. All patients that dropped out of the trial before week 5 were removed from the analysis (n=1283, 25.4%). For each patient, the total change in HAMD was determined and subsequently the contribution of each individual item to this change was calculated.

Box-plots were used to compare the contribution of each item between the different mechanisms of action. All graphical analyses and data manipulation were performed in the language and environment for statistical computing R (R Development Core Team, 2007).

Table 1. Characteristics of the included studies. For unpublished studies, see GlaxoSmithKline's clinical trial register (http://ctr.gsk.co.uk). Study 7 only included elderly patients

	no. pat.	Active treatments (dose)	Visits (T=titration design)	HAMD at baseline	Reference (NP=not published)
1	726	paroxetine (max 50 mg) imipramine (max 275mg)	1,2,3,4 6(T)	≥18	Feighner *et al.*
2	474	paroxetine (10 mg) paroxetine (20 mg) paroxetine (30 mg) paroxetine (40 mg)	1,2,3,4 6,9,12	≥18	Dunner and Dunbar
3	691	paroxetine (max 50 mg) fluoxetine (max 80 mg)	1,2,3,4, 6,9,12(T)	≥18	protocol 115 (NP)
4	848	paroxetine (max 50 mg) fluoxetine (max 80 mg)	1,2,3,4 6,9,12(T)	≥18	protocol 128 (NP)
5	315	paroxetine IR (max 50 mg) paroxetine CR (max 62.5 mg)	1,2,3,4 6,8,12(T)	≥20	Golden *et al.* Golden
6	330	paroxetine IR (max 50 mg) paroxetine CR (max 62.5 mg)	1,2,3,4 6,8,12(T)	≥20	Golden *et al.* Golden
7	319	paroxetine IR (max 40 mg) paroxetine CR (max 50 mg)	1,2,3,4,6 8,10,12(T)	≥18	Rapaport *et al.*
8	447	paroxetine CR (12.5 mg) paroxetine CR (25 mg)	1,2,3,4 6,8	≥20	Trivedi *et al.*
9	453	desipramine (max 200 mg) lamotrigine (max 200 mg)	1,2,3,4 6,7,8(T)	≥20	protocol 2011 (NP)
10	152	lamotrigine (max 200 mg)	1,2,3,4 5,6,7(T)	≥20	protocol 20022 (NP)
11	301	lamotrigine (max 200 mg)	1,2,3,4 5,6,7(T)	≥20	protocol 20025 (NP)

RESULTS

Clinical data

The percentage of patients remaining is summarised by treatment class and by week in figure 1. Only the first four weeks are shown because the time at which measurements were performed diverges between the studies after 6 weeks, which makes a comparison difficult. The number of patients in the dataset with measurements after week 4 was clustered, since this subset is later used in one of the analyses. Further evidence of the robustness of the data included in the final analysis is provided by the percentage of patients classified as responders (based on all measurements from each patient) at week 1, and in the dataset with patients remaining after week 4 (figure 2).

Item analysis

The time course of the distribution of the scores in responders and non-responders for *depressed mood* and *suicide* is shown in figure 3, separately for placebo and each of the mechanisms of action. These items were previously identified as sensitive items in the HAM-D_7 subscale (chapter 3). Sensitivity in this context is defined as the capacity to distinguish between responders and non-responders. No differences between the mechanisms of action were observed for any of seven items although fewer low scores were observed when considering the time course for responders to lamotrigine, as compared to the other mechanism of actions. Figure 4 depicts the time course for *loss of weight*

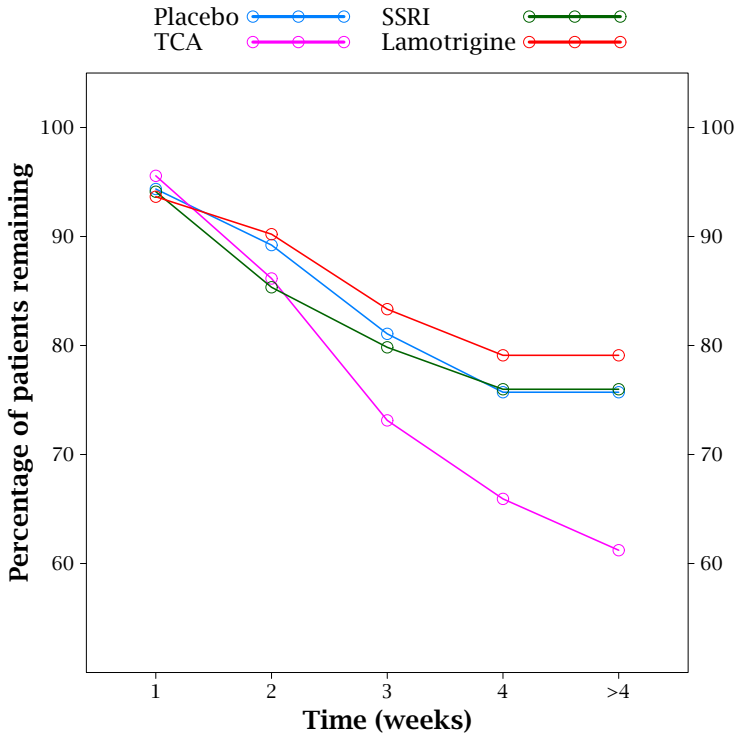
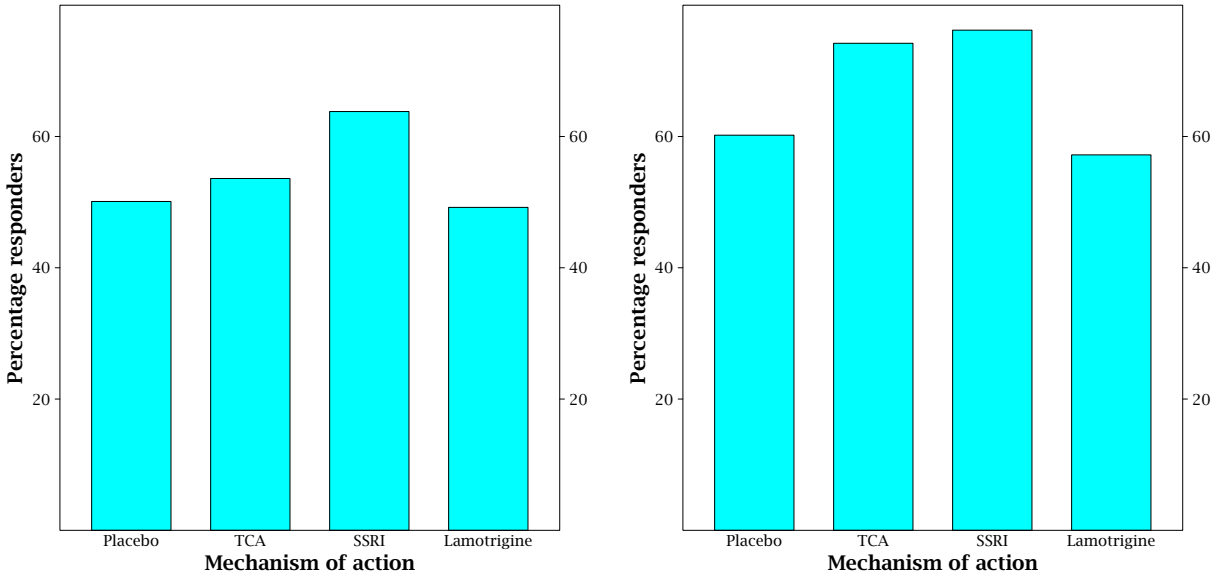
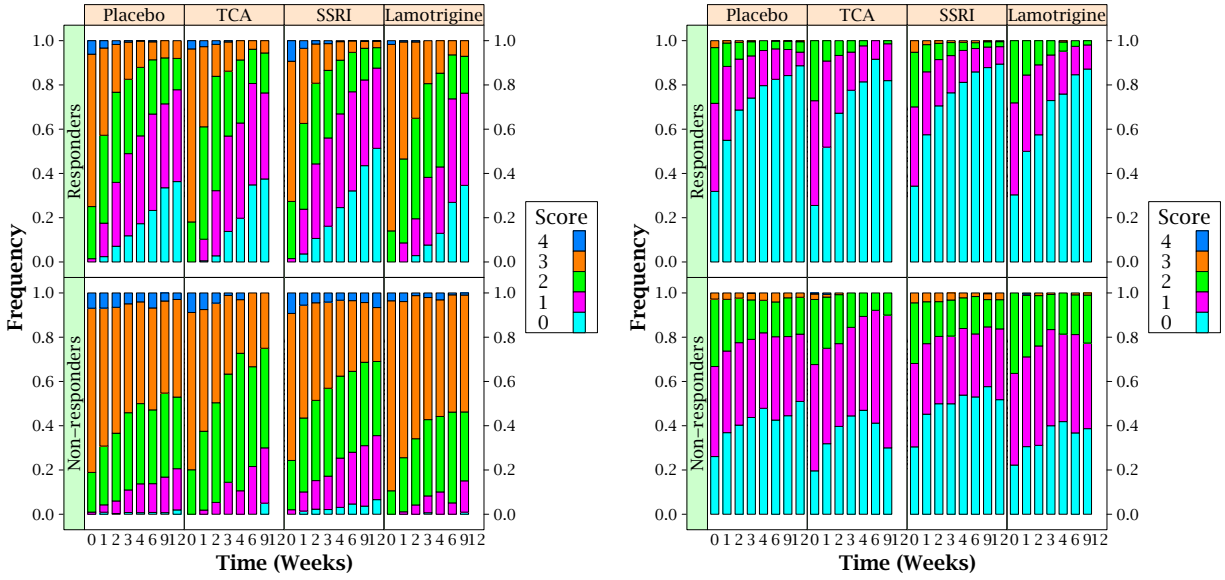


Figure 1. Percentage of patients remaining in the studies for each mechanism of action



(a) Week 1 (b) Patients with observations after week 4

Figure 2. Percentage of patients classified as responder based on all data

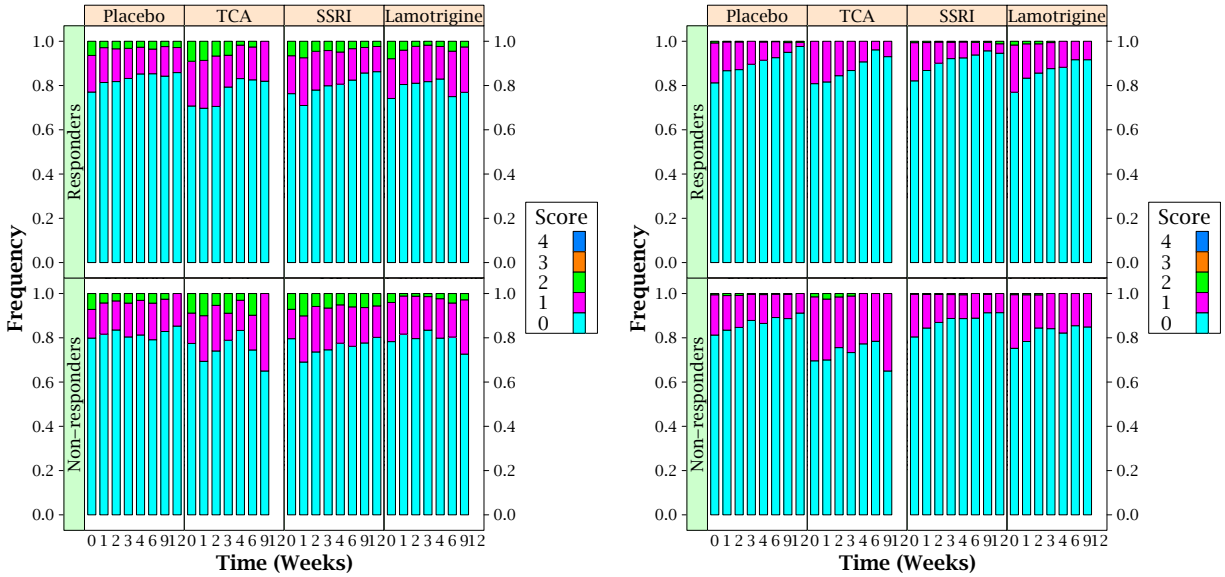


(a) *depressed mood* (b) *suicide*

Figure 3. Time course (in weeks) and score distribution for two response-sensitive HAMD items, separated by responders (upper panels) *versus* non-responders (lower panels) and mechanism of action (placebo, TCA, SSRI and lamotrigine)

and *loss of insight*. These two items were previously identified as insensitive to response. Clearly, there are no differences in their time course with respect to mechanism of action.

Evidence of differential effects on the extent of response at completion of treatment can be obtained by assessing the relative contribution of each item to the total change from baseline in HAMD at the last visit.



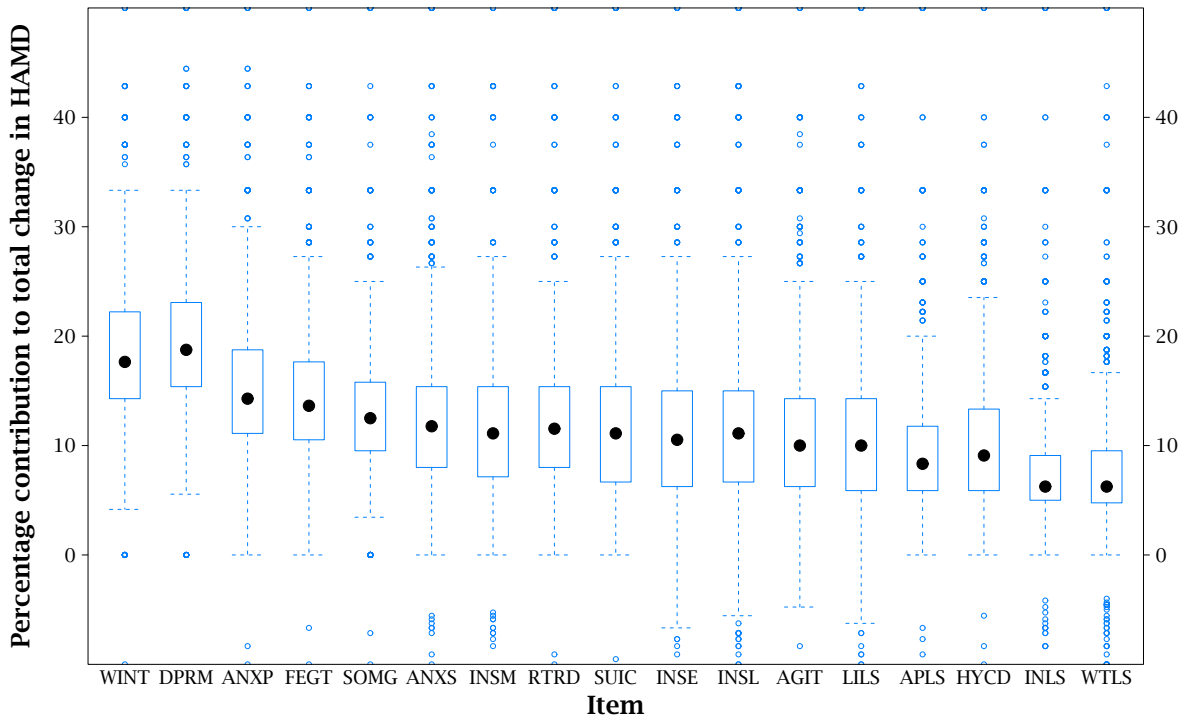
(a) *loss of weight* (b) *loss of insight*

Figure 4. Time course (in weeks) and score distribution for two response-insensitive HAMD items, separated by responders (upper panels) *versus* non-responders (lower panels) and mechanism of action (placebo, TCA, SSRI and lamotrigine)

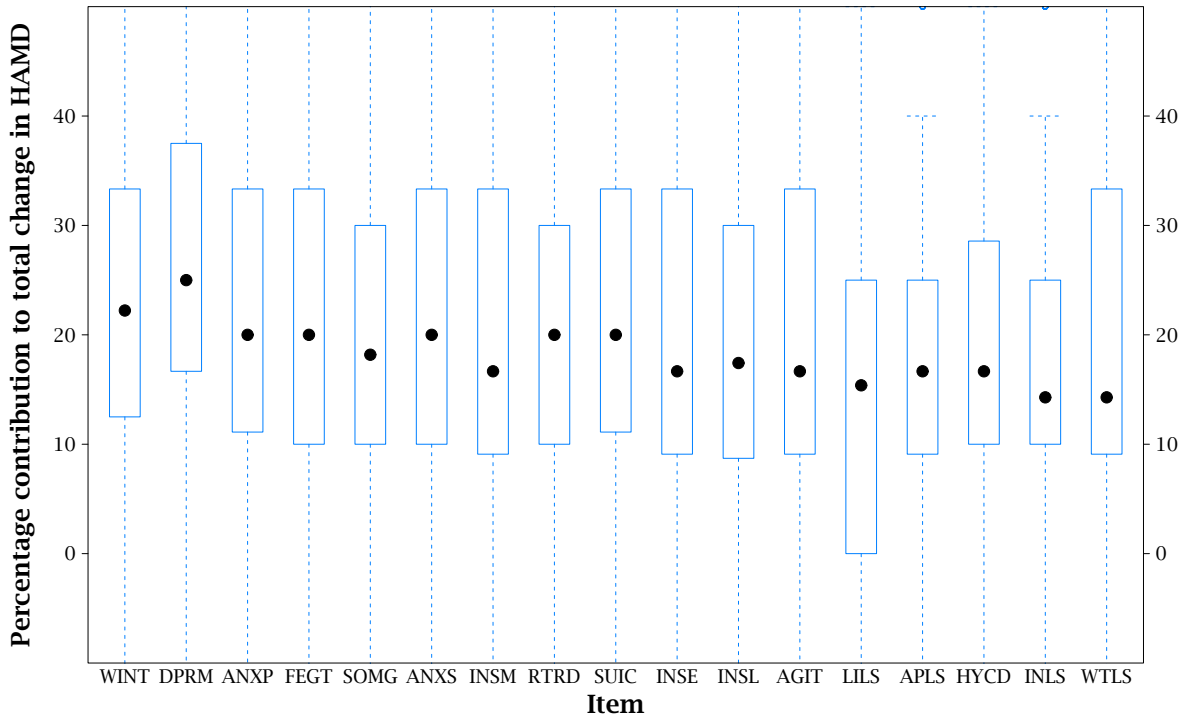
This is illustrated in figure 5 for responders and non-responders treated with SSRIs. As expected, the items considered sensitive to response, which are present in most HAMD subscales, contribute most to the total change in HAMD at completion of treatment. Another aspect of interest is that the variability of the contribution of each item for the total change in HAMD is much higher in non-responders than in responders, suggesting that the observed temporal patterns in responders during the course of treatment (approach 1) are specific throughout the course of therapy.

To allow for comparison between the different mechanisms of action, box-plots of the contributions of each item to the total change in HAMD were produced in separate panels. Figure 6 shows these findings for responders. No evidence is seen here that points to a class-effect of any of the items of the HAM-D_{17}. The order of the items reflects the contribution of the items to the total change in HAMD, with the most important ones being *work and interests* and *depressed mood*, followed by *psychic anxiety*, *feelings of guilt*, and *somatic symptoms general*. Then several items follow which are similar with respect to their contribution to the total change in HAMD. The least important items seem to be *loss of weight*, *loss of insight*, *loss of appetite* and *hypochondriasis*. Interestingly, the variability within treatment classes seems to be higher than the variability between classes. For non-responders, no clear order is observed amongst the items due to the high variability of the contributions of each item to the total change in HAMD.

For each item, an analysis of variance (ANOVA) was performed to assess whether the differences between the mechanisms of action in terms of their contribution to the total change in HAMD was statistically significant. No significant differences were found.



(a) responders



(b) non-responders

Figure 5. Box-plots of the contribution of each item to the total change in HAMD at the last visit for SSRI patients only. AGIT=*agitation*, ANXP=*anxiety psychic*, ANXS=*anxiety somatic*, APLS=*loss of appetite*, DPRM=*depressed mood*, FEGT=*feelings of guilt*, HYCD=*hypochondriasis*, INLS=*loss of insight*, INSE=*insomnia early*, INSL=*insomnia late*, INSM=*insomnia middle*, LILS=*loss of libido*, RTRD=*retardation*, SOMG=*somatic symptoms general*, SUIC=*suicidal thoughts*, WINT=*work and interests*, WTLS=*loss of weight*

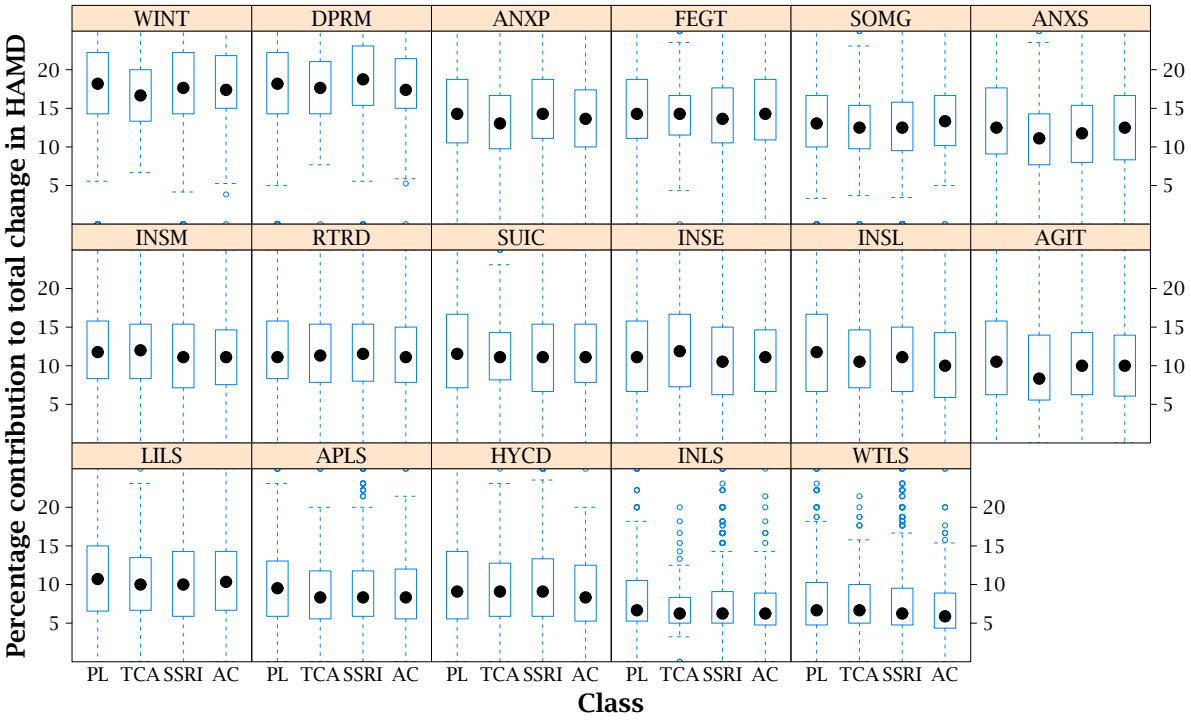


Figure 6. Contribution of each item to the total change in HAMD at the last observation by mechanisms of action. The items are ordered by decreasing average contribution. Only responders are shown. AGIT=*agitation*, ANXP=*anxiety psychic*, ANXS=*anxiety somatic*, APLS=*loss of appetite*, DPRM=*depressed mood*, FEGT=*feelings of guilt*, HYCD=*hypochondriasis*, INLS=*loss of insight*, INSE=*insomnia early*, INSL=*insomnia late*, INSM=*insomnia middle*, LILS=*loss of libido*, RTRD=*retardation*, SOMG=*somatic symptoms general*, SUIC=*suicidal thoughts*, WINT=*work and interests*, WTLS=*loss of weight*

DISCUSSION

Attempts to differentiate compounds, tailoring treatment to suit the needs of a heterogeneous group of patients who are currently diagnosed with depression, rely on the sensitivity of the unit of clinical measure to capture such differences in pharmacological properties. Often in biomarker validation research, reference is made to a requirement for clinically relevant measures that separate disease from drug specific properties. Ideally, a disease specific endpoint ensures a clear readout of response, and partly corroborates the validity of the measurement tool. Drug-specific endpoints may lead to bias, false positive and false negatives in the evaluation of response. However, such a situation poses a challenge to the identification of better targets and differentiation between compounds during development.

Our results show that sensitive items of the HAMD scale are not specific to any of the mechanisms of action under evaluation in the available clinical studies. The time course of the items illustrates that the pattern of response and non-response does not seem to differ between the MoA. The lack of specificity of the HAMD scale is confirmed by the investigation into the contribution of each individual item to the total change in HAMD. The obtained estimates for central tendency and dispersion are indistinguishable from

each other. The only trend observed in this analysis regards lamotrigine. We consider such a trend an artefact of the trial data, given that all lamotrigine trials included in this analysis failed to show statistically significant differences between placebo and active drug. In these circumstances, a reduction in the extent of response in responders may be expected, as compared to the other trials in which the active treatment can be separated from placebo.

An interesting although not unexpected finding is the difference between responders and non-responders. Where a clear pattern emerges for the contribution of the individual items on total change in HAMD for responders, the contribution of the separate items to the total change in HAMD in the non-responders is much more variable (figure 5). Since the same items are important contributors to the total change in HAMD across the different mechanisms of action, this is additional evidence that items can be distinguished based on their sensitivity to response, irrespective of treatment. The items that on average contribute most to the total change in HAMD are the same items that are often grouped in subscales of the HAMD, as for example the HAM-D_7 subscale developed according to the same graphical methodology used for the current work (chapter 3). In this subscale, the five items with the most contribution to the change of total HAMD are included (figure 6), plus *retardation* and *suicide*, which follow closely after these items together with some other items which have an approximately equal contribution.

Whilst one might expect specific changes in responders, patients that do not respond do not show any specific tendency in individual sensitive items (i.e., non-specific changes). This is striking if one considers that pharmacological differences exist in terms of potency, intrinsic activity and selectivity for the various receptor-subtypes. In contrast to the lack of selectivity of effects on the HAMD, these same drugs do show differential response based on other clinical measures, including markers of safety and tolerability. Early evidence of differential effect was shown in 1974 in cerebrospinal biomarkers (fluid metabolites of serotonin and noradrenaline) (Bertilsson *et al.*, 1974). On a clinical level these differential effects are shown by the componential approach used by Katz *et al.* (2004b,a). Their work reveals that differences in mechanisms of actions have differential effects on specific aspects of depression, and that the timing of these effects also differs between classes of drugs. Furthermore, a recent investigation has concluded that the loudness dependence of auditory evoked potentials, which is a measure for the central activity of the serotonergic system, can be used as a predictor of response to different classes of antidepressants (Juckel *et al.*, 2007).

The absence of any specific fingerprint for differences in pharmacological properties suggests that the HAMD reflects the outcome of a common pathway for these mechanisms. Although quantitative EEG study has shown that there are differences between placebo- and active-treatment responders (Leuchter *et al.*, 2002), a PET-study has found that the same regions change upon placebo and fluoxetine response, with fluoxetine responders exhibiting additional changes (Mayberg *et al.*, 2000). It is conceivable that these additional changes may cause the differences in HAMD score between patients treated

with fluoxetine and those in the placebo arm. Future imaging work should try to correlate the changes in the images to the changes in specific domains of depression to pharmacological properties such as receptor occupancy, as measured by the multicomponential method developed by Katz *et al.* (2004b).

Limitations

The analysis presented in this paper consists of 11 studies. Some important design characteristics were the same (placebo-controlled, randomised, patients with major depression), but others were different. Among these are the times at which the HAMD was administered, type of dosing (fixed dose/titration) and study population (adults, elderly). We chose not to be restrictive in this matter, allowing the inclusion of as much data as possible into this investigation. Since the elderly population constitutes only a minor fraction of the total patients in this investigation any discrepancies in response and disease characteristics will have little consequences for the results. The difference between dose titration versus fixed dose designs has no effect for the second methodology presented here since this includes only the last observation of each patient. With respect to the time course of the items some effect is expected but this will be quantitatively rather than qualitatively and should have no bearing on the conclusions.

Comparison to previous reports

Other authors have also investigated the possibility that the HAMD behaves specifically towards mechanisms of action. Nelson *et al.* (2005b) investigated the residual symptoms of treatment with fluoxetine (SSRI) and ruboxetine (a norepinephrine reuptake inhibitor). Their investigation included data from 2 studies with a total of 421 patients. Unfortunately, these studies were not placebo-controlled and only responders were included in the final analysis. The only difference that was found between fluoxetine and ruboxetine was that the decrease in sexual interest was larger for the patients treated with fluoxetine. The authors have also examined the effect size of each individual item in the same dataset (Nelson *et al.*, 2005a). This analysis also fails to show a difference between the two mechanisms of action. It is unfortunate that no distinction was made between responders and non-responders in the latter analysis. Interestingly, the items with the highest effect size largely correspond to those selected in our previous work (chapter 3), including the *suicide* item.

The primary objective of another analysis by Khan *et al.* (2002a, 2004) was to investigate the differences in effect size between the HAMD, MADRS and the clinical global impression - severity scale (CGI-S), but an additional hypothesis was that the HAMD would be better suitable to pick up the effects of TCA treatment than SSRI treatment. The report includes 208 patients from 11 trials in a single centre. Based on the observation that the effect sizes are similar across all endpoints for each mechanism of action they conclude that the HAMD is not biased towards TCAs. Therefore, their approach uses the other endpoints (MADRS and CGI-S) as external validation. It is conceivable however that these

endpoints are also biased to TCAs, which could influence their results. Because only 2 studies included the MADRS as clinical endpoint, we were not able to test the behaviour of the individual items of the MADRS across different mechanisms of action. However, since the MADRS is a global disease severity measure like the HAMD we anticipate a similar result.

Two studies by Moller *et al.* (2000, 1998) investigated the possible bias of the HAMD towards TCAs by comparing the percentage of responders based on the HAMD and on the Bech 6-item subscale of the HAMD. The conclusions of this work are that the HAMD is more sensitive to detect the drug effect of TCAs than SSRIs. This is explained by the higher percentage of responders in the SSRI group when the Bech-scale is used instead of the HAMD, and lower in those patients treated with TCAs. The data used in these investigations comes from double blind, but not placebo controlled studies in which approximately 320 patients were included. Even if this relatively small sample was taken at face value and the small difference considered significant, the fact that HAM-D_{21} was used to perform a responder analysis makes it harder to compare to other studies in which HAM-D_{17} is most frequently used. Lastly, it would have been interesting to perform the same analysis including placebo patients.

Future prospects

Given the results from our analysis and the evidence from previous reports, it is apparent that the HAMD does not discriminate between mechanisms of action. As indicated above, the absence of any specific fingerprint for the HAMD suggests that it reflects the outcome of a common pathway. However, it is important to stress that in the development of antidepressant drugs with new mechanisms of action, it should not be assumed that all mechanisms share the common pathways currently encompassed in the HAMD. Therefore we recommend further clinical research into the effects of new targets to be based on the contribution of the individual items of the HAMD to total change. Graphical representations like figure 3 and 4 may be used to deduce a fingerprint from a new MoA and compare it to existing medication. Simply taking HAMD changes at face value may lead to both over- and underestimation of true antidepressant effect. In this respect it is of interest to define new scales to determine antidepressant effects. In rheumatology, the disease activity score (DAS) has been developed (van der Heijde *et al.*, 1993), which is a composite scale consisting of a biomarker, symptom counts and a patient assessment of disease activity using a visual analog scale (VAS). Similarly, in depression, Katz *et al.* (2004a) have defined response based not only on the HAMD, but also on the CGI-S and global assessment scale (GAS) (Endicott *et al.*, 1976). Other relevant descriptors of pharmacology, such as PET-imaging and a combination of biomarkers should however also be taken into account.

Advancements in the evaluation of antidepressant drugs require a new clinical research paradigm. Such a change demands review of current beliefs in clinical psychiatry. Psychiatrists must acknowledge that pharmacological mechanisms underlie imbalances of

the mind. The nature of such changes can be exemplified by the current use of dexamethasone suppression tests to differentiate between psychotic depression and non-psychotic depression (Nelson and Davis, 1997). A mechanism-based approach in depression research needs to be implemented that allows clinical interpretation of biomarkers, which are consistent and valid (Mossner *et al.*, 2007). Moreover, in order to detect more specific effects of antidepressants componential models should be used which assess multidimensionality of symptoms and signs, rather than relying on a single measure of the severity of disease (Katz *et al.*, 2004a,b). This would not only represent an opportunity to differentiate single compounds, but would also facilitate the evaluation of drug combination therapies, allowing intervention with different drugs with respect to effect and timing of effect.

REFERENCES

Bagby RM, Ryder AG, Schuller DR, and Marshall MB (2004) The Hamilton depression rating scale: Has the gold standard become a lead weight? *Am J Psychiatry* **161**:2163–2177.

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

Bech P and Rafaelsen OJ (1980) The use of rating-scales exemplified by a comparison of the Hamilton and the Bech-Rafaelsen melancholia scale. *Acta Psychiatr Scand* **62**:128–132.

Bertilsson L, Asberg M, and Thoren P (1974) Differential effect of chlorimipramine and nortriptyline on cerebrospinal-fluid metabolites of serotonin and noradrenaline in depression. *Eur J Clin Pharmacol* **7**:365–368.

Dunner DL and Dunbar GC (1992) Optimal dose regimen for paroxetine. *J Clin Psychiatry* **53**:21–26.

Endicott J, Spitzer RL, Fleiss JL, and Cohen J (1976) Global assessment scale - procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psychiatry* **33**:766–771.

Feighner J, Cohn J, Fabre LF J, Fieve R, Mendels J, Shrivastava R, and Dunbar G (1993) A study comparing paroxetine placebo and imipramine in depressed patients. *J Affect Disord* **28**:71–79.

Golden R (2003) Efficacy and tolerability of controlled-release paroxetine. *Psychopharmacol Bull* **37 Suppl 1**:176–186.

Golden RN, Nemeroff CB, McSorley P, Pitts CD, and Dube EM (2002) Efficacy and tolerability of controlled-release and immediate-release paroxetine in the treatment of depression. *J Clin Psychiatry* **63**:577–584.

Gosens R, Zaagsma J, Meurs H, and Halayko AJ (2006) Muscarinic receptor signaling in the pathophysiology of asthma and COPD. *Respir Res* **7**:73.

Green B (2003) Lamotrigine in mood disorders. *Curr Med Res Opin* **19**:272–277.

Hamilton M (1960) A rating scale for depression. *J Neurol Neurosurg Psychiatry* **23**:56–62.

van der Heijde DMFM, van 't Hof M, van Riel PLCM, and van de Putte LBA (1993) Development of a disease-activity score based on judgment in clinical practice by rheumatologists. *J Rheumatol* **20**:579–581.

Jonsson E (2004) Graphical display of population data. *PAGE* **13**.

Juckel G, Pogarell O, Augustin H, Mulert C, Muller-Siecheneder F, Frodl T, Mavrogiorgou P, and Hegerl U (2007) Differential prediction of first clinical response to serotonergic and noradrenergic antidepressants using the loudness dependence of auditory evoked potentials in patients with major depressive disorder. *J Clin Psychiatry* **68**:1206–1212.

Katz MM (1998) Need for a new paradigm for the clinical trials of antidepressants. *Neuropsychopharmacology* **19**:517–522.

Katz MM, Houston JP, Brannan S, Bowden CL, Berman N, Swann AC, and Frazer A (2004a) A multivantaged behavioural method for measuring onset and sequence of the clinical actions of antidepressants. *Int J Neuropsychopharmacol* **7**:471–479.

Katz MM, Tekell JL, Bowden CL, Brannan S, Houston JP, Berman N, and Frazer A (2004b) Onset and early behavioral effects of pharmacologically different antidepressants and placebo in depression. *Neuropsychopharmacology* **29**:566–579.

Kennedy SH, Eisfeld BS, Dickens SE, Bacchiochi JR, and Bagby RM (2000) Antidepressant-induced sexual dysfunction during treatment with moclobemide, paroxetine, sertraline, and venlafaxine. *J Clin Psychiatry* **61**:276–281.

Khan A, Brodhead AE, and Kolts RL (2004) Relative sensitivity of the Montgomery-Asberg depression rating scale, the Hamilton depression rating scale and the clinical global impressions rating scale in antidepressant clinical trials: a replication analysis. *Int Clin Psychopharmacol* **19**:157–160.

Khan A, Khan S, Shankles E, and Polissar N (2002a) Relative sensitivity of the Montgomery-Asberg depression rating scale, the Hamilton depression rating scale and the clinical global impressions rating scale in antidepressant clinical trials. *Int Clin Psychopharmacol* **17**:281–285.

Khan A, Leventhal RM, Khan SR, and Brown WA (2002b) Severity of depression and response to antidepressants and placebo: An analysis of the Food and Drug Administration database. *J Clin Psychopharmacol* **22**:40–45.

de Kloet R, de Rijk RH, and Meijer OC (2007) Therapy insight: is there an imbalanced response of mineralocorticoid and glucocorticoid receptors in depression? *Nat Clin Pract Endocrinol Metab* **3**:168–179.

Kramer MS, Winokur A, Kelsey J, Preskorn SH, Rothschild AJ, Snavely D, Ghosh K, Ball WA, Reines SA, Munjack D, Apter JT, Cunningham L, Kling M, Bari M, Getson A, and Lee Y (2004) Demonstration of the efficacy and safety of a novel substance P (NK1) receptor antagonist in major depression. *Neuropsychopharmacology* **29**:385–392.

Leuchter AF, Cook IA, Witte EA, Morgan M, and Abrams M (2002) Changes in brain function of depressed subjects during treatment with placebo. *Am J Psychiatry* **159**:122–129.

Mandema JW and Danhof M (1992) Electroencephalogram effect measures and relationships between pharmacokinetics and pharmacodynamics of centrally acting drugs. *Clin Pharmacokinet* **23**:191–215.

Mayberg HS, Brannan SK, Tekell JL, Silva JA, Mahurin RK, McGinnis S, and Jerabek PA (2000) Regional metabolic effects of fluoxetine in major depression: Serial changes and relationship to clinical response. *Biol Psychiatry* **48**:830–843.

Moller H (2001) Methodological aspects in the assessment of severity of depression by the Hamilton depression scale. *Eur Arch Psychiatry Clin Neurosci* **251 Suppl 2**:II13–II20.

Moller HJ, Gallinat J, Hegerl U, Arato M, Janka Z, Pflug B, and Bauer H (1998) Double-blind, multicenter comparative study of sertraline and amitriptyline in hospitalized patients with major depression. *Pharmacopsychiatry* **31**:170–177.

Moller HJ, Glaser K, Leverkus F, and Gobel C (2000) Double-blind, multicenter comparative study of sertraline versus amitriptyline in outpatients with major depression. *Pharmacopsychiatry* **33**:206–212.

Montgomery SA and Asberg M (1979) A new depression scale designed to be sensitive to change. *Br J Psychiatry* **134**:382–389.

Moret C (2003) Current therapy and future treatment strategies for depression (Paris, January 2003). *J Psychopharmacol* **17**:337–341.

Mossner R, Mikova O, Koutsilieri E, Saoud M, Ehlis AC, Mueller N, Fallgatter AJ, and Riederer P (2007) Consensus paper of the WFSBP task force on biological markers: Biological markers in depression. *World J Biol Psychiatry* **8**:141–174.

Nelson JC and Davis JM (1997) DST studies in psychotic depression: A meta-analysis. *Am J Psychiatry* **154**:1497–1503.

Nelson JC, Portera L, and Leon AC (2005a) Are there differences in the symptoms that respond to a selective serotonin or norepinephrine reuptake inhibitor? *Biol Psychiatry* **57**:1535–1542.

Nelson JC, Portera L, and Leon AC (2005b) Residual symptoms in depressed patients after treatment with fluoxetine or reboxetine. *J Clin Psychiatry* **66**:1409–1414.

Pacher P and Kecskemeti V (2004) Trends in the development of new antidepressants. Is there a light at the end of the tunnel? *Curr Med Chem* **11**:925–943.

Pani L, Pira L, and Marchese G (2007) Antipsychotic efficacy: Relationship to optimal D2-receptor occupancy. *Eur Psychiat* **22**:267–275.

Papakostas GI and Fava M (2006) A metaanalysis of clinical trials comparing moclobemide with selective serotonin reuptake inhibitors for the treatment of major depressive disorder. *Can J Psychiat-Rev Can Psychiat* **51**:783–790.

Papakostas GI and Fava M (2007a) A meta-analysis of clinical trials comparing milnacipran, a serotonin-norepinephrine reuptake inhibitor, with a selective serotonin reuptake inhibitor for the treatment of major depressive disorder. *Eur Neuropsychopharmacol* **17**:32–36.

Papakostas GI and Fava M (2007b) A meta-analysis of clinical trials comparing the serotonin (5HT)-2 receptor antagonists trazodone and nefazodone with selective serotonin reuptake inhibitors for the treatment of major depressive disorder. *Eur Psychiat* **22**:444–447.

R Development Core Team (2007) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Rapaport MH, Schneider LS, Dunner DL, Davies JT, and Pitts CD (2003) Efficacy of controlled-release paroxetine in the treatment of late-life depression. *J Clin Psychiatry* **64**:1065–1074.

Trivedi MH, Pigott TA, Perera P, Dillingham KE, Carfagno ML, and Pitts CD (2004) Effectiveness of low doses of paroxetine controlled release in the treatment of major depressive disorder. *J Clin Psychiatry* **65**:1356–1364.

Visser SAG, Wolters FLC, Van der Graaf PH, Peletier LA, and Danhof M (2003) Dose-dependent EEG effects of zolpidem provide evidence for GABA(A) receptor subtype selectivity in vivo. *J Pharmacol Exp Ther* **304**:1251–1257.

Volkow ND, Gur RC, Wang GJ, Fowler JS, Moberg PJ, Ding YS, Hitzemann R, Smith G, and Logan J (1998) Association between decline in brain dopamine activity with age and cognitive and motor impairment in healthy individuals. *Am J Psychiatry* **155**:344–349.