



Universiteit  
Leiden  
The Netherlands

## **To fail or not to fail : clinical trials in depression**

Sante, G.W.E.

### **Citation**

Sante, G. W. E. (2008, September 10). *To fail or not to fail : clinical trials in depression*. Retrieved from <https://hdl.handle.net/1887/13091>

Version: Corrected 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/1887/13091>

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

# Chapter

## 1

# Clinical trials in depression: Challenges and opportunities

## CONTENTS

<b>1</b>	<b>Background</b>	10
<b>2</b>	<b>Depression</b>	11
2.1	Subtypes of depression	12
2.2	Aetiology of depression	13
<b>3</b>	<b>Treatment of depression</b>	13
3.1	The placebo effect	14
3.2	Tricyclic antidepressants	15
3.3	Monoamine oxidase inhibitors	15
3.4	Monoamine theory of depression	16
3.5	Serotonin-specific re-uptake inhibitors	17
3.6	Other developed mechanisms of action	18
3.7	Mechanisms of action in development	18
<b>4</b>	<b>Assessing the severity of depression</b>	19
4.1	The Hamilton depression rating scale	19
4.2	The Montgomery-Asberg depression rating scale	20
4.3	The componential approach	20
4.4	Imaging techniques	21
<b>5</b>	<b>Aspects of clinical trial design in depression</b>	22
5.1	Dropout	24
5.2	Clinical trial simulation	24
<b>6</b>	<b>Statistical analysis of depression endpoints</b>	25
6.1	LOCF change from baseline	25
6.2	Mixed model for repeated measures	26
6.3	Hierarchical linear model	26
6.4	Survival analysis	26
6.5	Pharmacokinetic-pharmacodynamic modelling	27
6.6	Functional data analysis	28
6.7	Bayesian statistics	28

## 1 BACKGROUND

The design and evaluation of clinical trials with antidepressant drugs still constitutes a major challenge. An analysis of the Food and Drug Administration (FDA)-database has shown that even for known effective and marketed antidepressants up to 50% of trials had failed to show a statistically significant drug effect (Khan *et al.*, 2002b). Since the FDA currently demands two positive pivotal clinical trials before registration is considered (although a single pivotal trial in combination with confirmatory evidence also suffices), often 3-5 trials have to be performed due to the high failure rate. Undoubtedly, the factors contributing to this high failure rate are many. These factors may be divided into three main classes: disease, drug and trial design-related factors. Among the first are the high variability in response, the heterogeneity of patients being diagnosed with major depressive disorder (MDD), the difficulties in objectively measuring the severity of depression and the high placebo effect. Drug-related factors include difficulties in dose selection in the absence of relevant concentration-effect relationships from preclinical models, pharmacokinetic variability and poor compliance to treatment due to side effects. Trial design-related factors comprise inadequately sized study populations, sub-optimal inclusion criteria, insufficiently long durations of trials, use of insensitive endpoints rather than readily available more sensitive endpoints and suboptimal statistical analysis methods for the assessment of treatment effect.

The disease-related causes are generally hard to address in individual trials. For example, a possible solution to patient heterogeneity would be to redefine diagnostic criteria as to ensure more homogeneous groups of patients. Such a circumstance would demand changes in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (American Psychiatric Association, 1994) criteria and require revisiting the current perception of mood disorders as a whole. In contrast, trial design-related factors can be more readily addressed. The variability in response for example has often been assigned to the placebo effect. In fact, changes in clinical trial design have been suggested to mitigate this effect. Single- and double-blind placebo run-in phases have been introduced to detect placebo-responders early in the trial, albeit with limited effectiveness (Faries *et al.*, 2001; Lee *et al.*, 2004; Trivedi and Rush, 1994). The difficulties in assessing depression severity and the known limitations of clinical rating scales may be solved by developing multidimensional endpoints, or alternatively, by allowing for the introduction of composite measures in which additional objective (bio)markers of disease could be used as endpoints.

Failure due to drug-related factors, such as an inadequate dose selection, may be reduced by using a rationale, which takes into account pharmacokinetic-pharmacodynamic (PKPD) relations (Yassen *et al.*, 2007). Often dose selection in depression trials is still based on the maximum tolerated dose (MTD) approach, irrespective of whether evidence exists that such a dose is appropriate to achieve an optimal drug effect. The development of imaging techniques such as positron emission tomography (PET) enables characterisation of a binding profile, providing a proof of mechanism for the compound. If linked to

clinical response under steady-state conditions, this type of data can guide dose selection and can be used to explain some of the variability in response due to differences in drug exposure at the biophase. Another relevant factor associated with drug exposure is patient compliance. Poor compliance may lead to variable or insufficient clinical response, which consequently results in lack of separation from the placebo effect. Technologies to accurately monitor compliance are available, but the impact of compliance on outcome is often ignored in clinical protocols. Considering the difficulties mentioned above, one does not need to stress the importance of optimising clinical trial design factors as much as possible, as to prevent failure of clinical trials in depression for reasons other than true inefficacy of drugs.

The pressure to overcome these hurdles is high since antidepressant drugs are being developed by all major pharmaceutical companies. The need for novel antidepressants is evident from the non-response rate of about 30%, which is observed for currently approved drugs, and also from the multitude of side effects experienced by the target population.

The high failure rate of clinical trials has important consequences. Patients randomised to placebo in studies which fail to show a significant treatment effect due to a false negative result are exposed to an ineffective treatment without accomplishing the ultimate goal of clinical research, i.e., providing evidence of (absence of) benefit for the patient population. Also, clinical development plans for antidepressant drugs suffer considerable delays due to such negative trials. In the most extreme case, the development of efficacious drugs may be stopped, costing billions to the pharmaceutical industry and most importantly, depriving depressed patients from better medication.

Various aspects of clinical trials with antidepressants will be further discussed in the next sections. Section 2 will focus on depression itself, whilst section 3 will provide an overview of the available treatment options. Section 4 will discuss the methods currently available to assess depression severity, whereas section 5 will highlight some other important aspects of clinical trial design in depression. In section 6, we conclude with a review of the statistical analysis methods suitable for the evaluation of efficacy in clinical trials in depression.

## 2 DEPRESSION

The World Health Organisation (WHO) defines depression as "a common mental disorder that presents with depressed mood, loss of interest or pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, low energy, and poor concentration." It continues to say that "These problems can become chronic or recurrent and lead to substantial impairments in an individual's ability to take care of his or her every day responsibilities. At its worst, depression can lead to suicide, a tragic fatality associated with the loss of about 850,000 lives every year" (WHO, 2007).

About 121 million people suffer from depression worldwide. The National Institute of Mental Health (NIMH) reports a year-prevalence of 9.5%, corresponding to 18.8 million Americans (National Institute of Mental Health, 2000). Its Dutch counterpart, the Rijksinstituut voor Volksgezondheid en Milieu (RIVM), reports a year-prevalence of 6.3%, with females having twice the prevalence of males (Schoemaker *et al.*, 2005). Beside the staggering number of depressed patients, their relatives, friends and society in general also suffer its consequences. A widely used measure to compare the burden of diseases, the 'disease adjusted life years (DALYs)', ranks depression on the 4th place. It is projected that depression will reach the 2<sup>nd</sup> place in 2020 (WHO, 2007).

The scientific community is well aware that depression is an important disease. A PubMed search on the MeSH keyword 'depressive disorder' returns over 50,000 hits, with 5,500 dating from the last year. Furthermore, over 1,000 active clinical trials have been registered at [www.clinicaltrials.gov](http://www.clinicaltrials.gov) with the keyword 'depression'. Given the high occurrence and social profile of depression today, it is perhaps surprising to note that the disease was recognised only around the start of 20<sup>th</sup> century (Kraepelin, 1896), although others suggest that it was recognised as early as the 2<sup>nd</sup> century AD (Sartorius, 2001).

## 2.1 Subtypes of depression

The first and second edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM) (American Psychiatric Association, 1952, 1968) held a dimensional view on psychiatric disorders. The leading hypothesis at that time was the existence of a continuum between patients and healthy subjects. Partly because of the development of 'specific' drugs against psychiatric symptoms, a more categorical view has gained popularity since the introduction of the third edition of the DSM (American Psychiatric Association, 1980; Healy, 1997). Depression, like mania, is one of the affective disorders, i.e., disorders that influence the mood of the patient in this third edition. Often periods of mania are alternated by periods of depression, leading to the term 'bipolar disorder' or 'bipolar depression'. Several types of depression have been identified, for example mild depressive disorder, moderately severe depressive disorder and major depressive disorder (MDD) (American Psychiatric Association, 1994). The most important symptoms of mild depressive disorder include low mood, disturbed sleep and lack of energy and enjoyment. In moderately severe depressive disorder the patient's appearance may also change, known as a state of psychomotor retardation. In MDD, the severity of all of these symptoms is increased and some patients may additionally experience delusions (psychotic depression) (American Psychiatric Association, 1994). The latter category is usually excluded from clinical trials in depression. Other types of depression include seasonal affective disorder (SAD), in which patients experience depression only in a particular time of year (mostly winter).

Depression may manifest itself at any stage in life, most first episodes however are identified in patients in the age of 25-40 (Weissman *et al.*, 1996). The median duration of an episode is 3 months, although 20% of depressed patients may have episode durations of over 2 years (Spijker *et al.*, 2002). Over 75% of the patients who experience a single

episode will have recurrent episodes in the following ten years (Piccinelli and Wilkinson, 1994).

It is important to note that in current clinical practice, the diagnosis of depression is based primarily on symptoms, neglecting the potential differences in the aetiology of the disease. Patients with different underlying pathologies may therefore be grouped under the same diagnosis, contributing to the heterogeneity between patients.

## 2.2 Aetiology of depression

Much of the aetiology of depression is as yet unknown. Only pharmacology-related theories will be discussed in the scope of this thesis. An important hypothesis derived from the efficacy of a class of drugs, is the monoamine theory (see section 3.4). It is believed that an imbalance in monoamines, especially serotonin and norepinephrine, may predispose for depression in later life. Often, a precipitating event can be identified as a direct cause for the first depressive episode. These are usually stressful life events, such as the death of a close relative or friend, loss of employment, *et cetera*. Also, depression often occurs in combination with other illnesses, such as Parkinson's disease (Lieberman, 2006). Further details on the pathophysiological nature of depression, are presented in section 3.4.

## 3 TREATMENT OF DEPRESSION

Many options are available to treat patients suffering from depression. Unfortunately, more than 30% of patients will become resistant to treatment in the course of therapy (Fava and Davidson, 1996). Frequently used treatments include cognitive behavioural therapy, pharmacological interventions and electro-convulsive therapy (ECT). Other, less well evidenced treatments, such as herbal medicine (e.g., St John's wort, the most frequently used antidepressant treatment in Germany) and other alternative therapies are also used. The Cochrane institute has withdrawn plans for a review comparing the efficacy of psychotherapy and pharmacological therapy and the benefits of a combination of the two. Another review, however, found that psychotherapy enhances the effects of pharmacotherapy, partly due to improving compliance (Pampallona *et al.*, 2004). The next sections will discuss the pharmacological treatment options in more detail.

It is important to take the history of the development of antidepressants into account to understand the current practice in this field. This history has been extensively described by Healy (1997), and will be summarised in the next paragraphs.

### 3.1 The placebo effect

Currently, the most accepted definition of placebo effect is the one which is due to the meaning of a therapeutic intervention for a particular patient (Moerman and Jonas, 2002). Throughout this thesis we will adopt the widespread use of 'placebo effect' as the treatment effect demonstrated in the placebo arm of a trial. In this connotation, 'placebo effect' will also encompass the natural history of the disease, as depression episodes tend to end spontaneously.

Although placebo may not be the first thought when thinking of a pharmacological treatment, it is the only intervention present in most clinical trials across a wide range of diseases. In fact, the placebo effect is a very important factor in drug development if one considers that placebo response rates vary from 12.5-51.8% (on average 29.7%) in clinical trials in depression (Walsh *et al.*, 2002). Several clinical trial design features have been suggested and applied in order to minimise placebo response rates. These are discussed in section 5.

An interesting aspect of the placebo effect is that it is growing over time (Walsh *et al.*, 2002). As discussed elsewhere, the most probable causes for this time trend are a change in the type of patients that is included in clinical trials (in-patient *versus* out-patient), a decreased disease severity in the included patients and the widespread use of and knowledge about antidepressants which enhance the patient's expectation about the effectiveness of pharmacological intervention (Walsh *et al.*, 2002). An important difference between placebo and 'active' pharmacological treatment is the absence of drug-specific side effects. It has been hypothesised that despite the high placebo response rate in depression, the presence of well-known side effects, which are often stated in the informed-consent forms that are signed by each patient, may unblind patients to the treatment they are receiving and inflate the efficacy of antidepressants. Indeed, a recent review by the Cochrane group revealed that the effect size of antidepressants when compared to active placebo (a placebo with side effects similar to those observed in antidepressants) greatly diminished, although it remained statistically significant (Moncrieff *et al.*, 2004).

The use of imaging techniques (see section 4.4) may shed more light on the nature of the placebo effect, and may allow a separation of placebo-responders from non-responders, allowing for a better estimate of the effect of active treatments (Vallance, 2007). However, due to the complicated interaction between placebo and drug effect, as clearly conceptualised by Moerman and Jonas (2002), it may never be possible to separate these two effects entirely.

An important conclusion that can be drawn from the above is that the variability of the placebo effect does place doubt upon studies which are not placebo controlled (Khan *et al.*, 2002a; Walsh *et al.*, 2002). Furthermore, the difficulties caused by the placebo effect are increased by the choice of insensitive endpoints, sub-optimal doses and other trial-related factors.

### 3.2 Tricyclic antidepressants

Contrary to common misconception, the tricyclic antidepressant (TCA) imipramine was the first antidepressant to be discovered (Healy, 1997). In 1950 a structural analogue of imipramine failed to show effect on a range of psychiatric diseases. Only after the clinical effects of the anti-psychotic chlorpromazine were demonstrated, the psychiatric world considered the possibility of drugs that were effective against psychiatric illnesses. This renewed the interest in structural analogues of imipramine, and the analogue most closely resembling chlorpromazine was subsequently tested as an anti-psychotic. This trial failed, but interesting side effects relating to mood elevations were found and in 1955 imipramine became the first drug to be tested in depressed patients (Kuhn, 1957, 1958). The first patient responded after as little as 6 days. However, it took further priming of the psychiatric community and the development of another class of antidepressants, the monoamine oxidase inhibitors (see next section), for imipramine to become an established antidepressant in 1960 (Rees, 1960).

TCAs are thought to exert their antidepressant effect by inhibiting the presynaptic re-uptake of serotonin (5-HT) and norepinephrine, increasing the concentrations of these neurotransmitters to restore the balance in monoamine homeostasis (Feighner, 1999; Stahl, 1998).

The most important side effects of TCAs are dry mouth, constipation, weight gain, sexual dysfunction and cardiovascular disturbances (at high exposure levels). Interestingly, it is believed that the effect of TCAs manifests itself with a delay of two to six weeks after the start of treatment (Feighner, 1999; Stahl, 1998), even though miraculous improvement within 6 days has been reported in the first trial (Kuhn, 1957) and earlier response is claimed by other investigations (Posternak and Zimmerman, 2005; Stassen and Angst, 1998). Until the late 1980s, TCAs were the drug of first choice in pharmacological depression therapy. Well-known TCAs are imipramine, amitriptyline and desipramine.

### 3.3 Monoamine oxidase inhibitors

Monoamine oxidase inhibitors (MAOIs) were discovered independently of and around the same time as the TCAs (Healy, 1997). The first MAOI, iproniazid, was an anti-tubercular drug. It was found that it raised mood levels in patients, which was first considered a side effect and caused the drug to be replaced. In 1956 a study was started to evaluate the properties of iproniazid as a 'psychic energiser' which led to the establishment of this drug as an antidepressant (Loomer *et al.*, 1957).

The mechanism of MAOIs is thought to be the inhibition of the oxidation of monoamines such as serotonin (5-HT), norepinephrine and dopamine, thereby indirectly raising the concentrations of these amines (Feighner, 1999; Stahl, 1998).

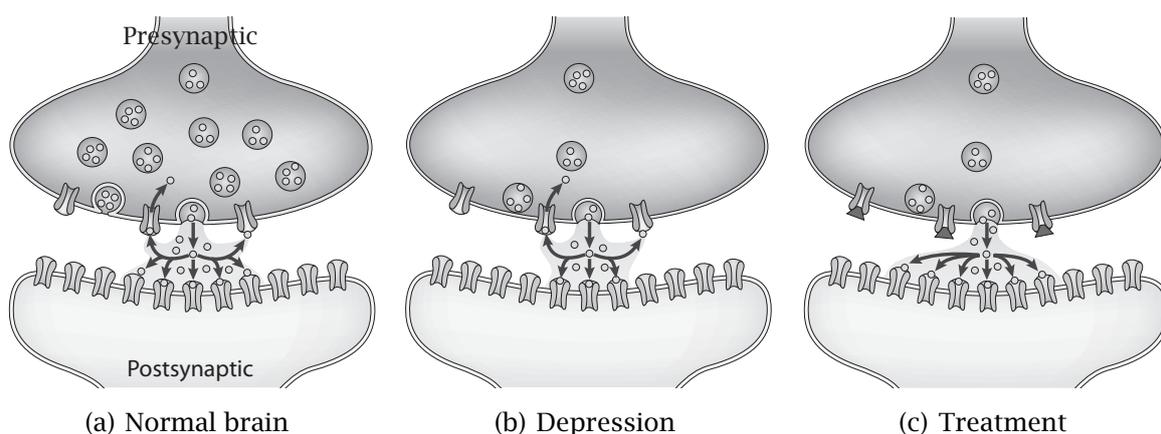
The most important side effect of MAOIs is the interaction with food containing significant amounts of tyramine, such as seasoned cheese and wine. Normally, amines contained in these foods are degraded by the monoamine oxidase enzyme, but this process is inhib-

ited in the presence of MAOIs. Tyramine may cause high-blood pressure and increased heart rate, eventually leading to intracranial bleeding. MAOIs are safe drugs if care is taken to prevent food interaction (Feighner, 1999; Stahl, 1998). The MAOIs currently on the market are phenelzine, tranylcypromine and isocarboxazid.

### 3.4 Monoamine theory of depression

The mechanisms of action of MAOIs and TCAs led to the monoamine theory of depression (Schildkraut, 1965). This theory hypothesises that depression is caused by a shortage of one or more neurotransmitters in the brain, serotonin and norepinephrine in particular (figure 1). Both MAOIs and TCAs cause an increase of the concentration of serotonin and/or norepinephrine in the synaptic cleft, counteracting this putative shortage. This theory led to the development of a new class of antidepressants, the serotonin specific re-uptake inhibitors (SSRIs). It was thought that the specificity of SSRIs, in contrast to TCAs, should decrease the side effects that were associated with TCA use without limiting efficacy. The main flaw in this theory is that it fails to explain the action of more recently developed antidepressants such as tianeptine, which increases serotonin re-uptake, and does not explain why antidepressants also work against, e.g., anxiety disorders (Hindmarch, 2002). Also, studies invoking monoamine depletion in healthy subjects have shown that this does not lead to a depressive syndrome, although monoamine depletion in depressive patients responding to treatment does lead to a reversal of the treatment effect, as discussed in Delgado (2000).

Although some, like Leonard (2000), maintain that the monoamines should be the primary focus in antidepressant research, the general consensus is that other mechanisms are likely to be important as well, such as effects on the hypothalamic-pituitary-thyroid axis (HPT) and the hypothalamic-pituitary-adrenal (HPA) axis (Hindmarch, 2002). The in-



**Figure 1.** Monoamine hypothesis of depression. (a) In the normal brain, monoamine neurotransmitters (smallest circles) are released and bind to receptors on the postsynaptic neuron. Transmission is terminated by re-uptake of the transmitter. (b) In depression, the decreased concentration of monoamine at synaptic sites produces a mood disorder. (c) Blockade of the re-uptake sites increases the concentration of monoamine neurotransmitters available at receptor sites and restores mood. Adapted with permission from Castren (2005)

involvement of the HPT was first suggested because patients with hypothyroidism display symptoms similar to those observed in depressed patients. Furthermore, many depressed patients have a reduced thyroid function (Jackson, 1998). Involvement of the HPA axis is supported by reports that cortisol treatment for patients with Cushing's syndrome improves psychiatric symptoms (Sonino *et al.*, 1993), and by studies showing a hyperactive HPA axis in many depressed patients (Holsboer *et al.*, 1995; Rubin *et al.*, 1995). It may be that the action of cortisol becomes harmful after increased periods of stress due to changes in the dynamics of the mineralocorticoid and glucocorticoid receptors, increasing the vulnerability to depression, such as hypothesised by de Kloet *et al.* (2007). Other components of the stress response system such as the hippocampus and the immune system are also thought to be involved. The hippocampus has a reduced volume in depressed patients (Bremner *et al.*, 2000; Sapolsky, 2000; Sheline *et al.*, 1996) and treatment with antidepressants increases the volume of the hippocampus (Drew and Hen, 2007).

None of the above mechanisms are able to fully explain the changes and consequences of depression. Some may play a role in early stages of disease, whereas others are associated with morbidity and relapse. An integrative approach, fully accounting for the observed disease progression, may shed more light on the interrelations between these mechanisms.

### 3.5 Serotonin-specific re-uptake inhibitors

The discovery of the first serotonin-specific re-uptake inhibitor (SSRI), fluoxetine (Prozac), was published in 1974 (Wong *et al.*, 1974). It was only in 1985 that its antidepressant effects were recognised (Cohn and Wilcox, 1985) and in 1987 it was finally registered as an antidepressant in the USA. SSRIs were developed based upon the monoamine theory and inhibit the re-uptake of serotonin. Side effects of SSRIs are similar to those observed upon treatment with TCAs, although their incidence is generally lower, as expected from the more specific mechanism (Steffens *et al.*, 1997). Another advantage of SSRIs is that the risks for adverse events associated with overdosing are lower compared to TCAs. On the other hand, SSRIs have been linked to a higher incidence of suicide compared to placebo treatment (Lenzer, 2006; Healy, 2006) and to increased side effects regarding sexual dysfunction, which differs significantly between the SSRI compounds (Westenberg and Sandner, 2006). In fact, this side effect may lead to a new indication (premature ejaculation) for SSRIs (Wang *et al.*, 2007).

SSRIs continue to be the first choice of pharmacological intervention in depression. Well-known SSRIs are fluoxetine, paroxetine and fluvoxamine. In 2005 in the Netherlands alone, 1.5 million paroxetine prescriptions were sold for a value of 52 million euros, making it the 7<sup>th</sup> item in the list of pharmaceuticals where most money was spent on (SFK, 2006).

### 3.6 Other developed mechanisms of action

Other mechanisms of action which have been the focus of drug discovery include combinations of serotonin re-uptake inhibition with re-uptake inhibition of other neurotransmitters. Examples are serotonin-norepinephrine re-uptake inhibitors (SNRIs), such as venlafaxine and duloxetine, norepinephrine-dopamine re-uptake inhibitors (bupropion) and norepinephrine re-uptake inhibitors (reboxetine). These drugs, which were developed based on the monoamine theory, have a similar efficacy and side effect profile compared to SSRIs (Stahl, 1998; Artigas *et al.*, 2002).

In parallel to the advancements of drug development, the use of herbal therapies has grown considerably. As mentioned earlier, a popular antidepressant in Germany is St. John's Wort (*Hypericum perforatum*). Hyperforin, one of the active ingredients, has been shown to inhibit the re-uptake of serotonin and norepinephrine directly and via an unknown indirect mechanism (Leuner *et al.*, 2007). Although a review by the Cochrane institute concluded that efficacy is doubtful (Linde *et al.*, 2005), recent studies have indicated that St John's Wort is better than placebo (Kasper *et al.*, 2006) and comparable to paroxetine (Anghelescu *et al.*, 2006) in the treatment of depression.

### 3.7 Mechanisms of action in development

Currently, various new mechanisms of action are being developed as targets for the treatment of depression. However, it is sightlessly assumed that the physiological function(s) associated with these targets are reversible and relevant for the course of depression at the timing of the intervention.

An integration of the mechanisms hypothesised to play a role in depression into a disease progression model may provide valuable insight into relevant targets and opportunities for treatment. This means revisiting current practice in drug discovery, allowing for an approach which focuses on the role of specific pathways, rather than on target identification only. In addition, interventions could be considered that treat a prodromic phase of disease. Prophylaxis may be meaningful for instance in specific subpopulations with disturbed stress-response systems.

Whilst anti-depressant efficacy in patients has already been demonstrated for some mechanisms, other drugs have yet to reach the clinical phase. The best-known new class are the neurokinin-1 (NK1) antagonists, which block the neurokinin (or substance P) receptor. These have been reported to have antidepressant activity in clinical trials (Kramer *et al.*, 2004), although evidence is mounting that the NK1-antagonists may not be as effective as expected (Czeh *et al.*, 2006). Hybrid compounds that block the neurokinin receptor as well as inhibit the re-uptake of serotonin are also being investigated.

Another promising class of drugs are the corticotropin-releasing factor (CRF) receptor antagonists (Nielsen, 2006). Some evidence suggests that CRF is hypersecreted in depression. In an open-label clinical trial a CRF-antagonist appeared to be effective (Zobel *et al.*, 2000), but development was later stopped due to adverse events at higher doses.

## 4 ASSESSING THE SEVERITY OF DEPRESSION

One of the main difficulties of drug development in depression is adequate and objective assessment of disease severity. This is an important aspect if one wants to evaluate the beneficial effect of compounds which provide symptomatic relief or show disease modifying properties. There have been propositions suggesting that the only meaningful outcome criteria are those which are completely objective, such as suicide rates or shortened hospital stays (Kline, 1959). However, in general it is accepted that it is necessary to make some assumptions and substitute truly clinical endpoints for surrogate endpoints such as biomarkers (Atkinson *et al.*, 2001). As an example of surrogate endpoints, several rating scales have been developed which aim at scoring the symptoms of depression as determined in the DSM-criteria (American Psychiatric Association, 1994). These rating scales may be divided into two categories, observer-rated scales and self-rated scales. The best-known observer-rated scales are the Hamilton depression rating scale (HAMD) (Hamilton, 1960) and the Montgomery-Asberg depression rating scale (MADRS) (Montgomery and Asberg, 1979). More general scales are also used, such as the clinical global impression (CGI) (Guy, 1976), which score the clinician's global impression of the patient rather than trying to capture the various aspects of the disease state. A well-known self-rating scale is the Zung self-rating depression scale (Zung, 1965).

The next sections will concentrate on various aspects of the observer-rated depression scales investigated throughout this thesis, the HAMD and MADRS, and discuss an alternative rating scale as well as the use of imaging (e.g., PET) techniques to assess depression severity. Self-rated subscales were not included since these were not consistently available in the historical data that was retrieved for our investigations.

### 4.1 The Hamilton depression rating scale

The Hamilton depression rating scale (HAMD), developed by Hamilton (1960), was the first broadly accepted depression rating scale. Its popularity may be partly explained by the simultaneous development of the TCAs, whose effects could be captured by the HAMD (Broadhurst and Healy, 1996). It is intended to be scored by a trained rater after an interview of approximately 30 minutes. In its original form, it consists of 17 items attempting to cover all areas that are affected by the disease. A patient with a score higher than 18 is considered depressed, a reduction of 50% or greater is considered response (Tedlow *et al.*, 1998), whereas a HAMD score of  $\leq 7$  is considered remission. The purpose of the scale was to determine the severity of depression once it had been diagnosed, but not as a tool to measure changes upon treatment administration (Hamilton, 1960).

Many modified versions of the scale have been developed over the years (Potts *et al.*, 1990; Miller *et al.*, 1985; Gelenberg *et al.*, 1990; Paykel, 1985; Hamilton, 1967), which has caused some problems because it is not always clear which depression scale has been used in a particular clinical trial (Williams, 2001). The original paper by Hamilton

established good between-rater consistency. Later, the multidimensionality of the HAMD was discovered and criticised (Bech and Rafaelsen, 1980; Moller, 2001). In the context of rating scales, multidimensionality means that more than one aspect of the disease is measured, and thus that two patients may have the same scores but different disease features. Another criticism on the HAMD is that it contains 3 items related to sleep (*early, middle & late insomnia*), which may cause the HAMD to be more sensitive towards the effect of TCAs than to SSRIs, because of the sedative properties of the TCAs (Moller, 2001). Additionally, TCAs may cause significant weight gain which might be reflected in the weight-related item, *loss of weight*.

A Rasch item-response analysis by Bech and Rafaelsen (1980) has led to the identification of several unidimensional subscales measuring only one aspect of depression. One of these subscales, the 6-item core-subscale of depression, is often advocated as a more sensitive endpoint to detect drug effect. Other subscales of the HAMD have also been suggested, by investigators such as Maier and Philipp (1985). Recently, an item response analysis revealed that not all items represent different disease levels (Evans *et al.*, 2004). These findings share a common conclusion, i.e., that subscales outperform the HAMD in the detection of treatment effect (O'Sullivan *et al.*, 1997; Faries *et al.*, 2000).

Despite the issues raised and solutions offered in the form of subscales, the HAMD remains the gold standard for clinical trials (Bagby *et al.*, 2004). Quicker and cheaper methods of applying the HAMD have been investigated, such as interactive voice response (IVR). This method has been shown to be highly correlated to a HAMD administered by a clinician (Moore *et al.*, 2006). Clinical trials have been successfully performed with this new method of administering the HAMD (Pierre and Kasper, 2007).

## 4.2 The Montgomery-Asberg depression rating scale

In 1979, 19 years after the HAMD was first published, the Montgomery-Asberg depression rating scale (MADRS) was developed with the explicit aim to be sensitive to change (Montgomery and Asberg, 1979). The MADRS consists of 11 items and covers a broad range of antidepressant symptoms. Because less emphasis is placed on insomnia-related items, it is supposed to be less specific to TCAs. Like the HAMD, it is intended to be completed by a trained rater after an interview of about half an hour. Recently, an IVR application of the MADRS has been validated by showing a high correlation between total IVR MADRS and the MADRS as rated by a clinician. In fact, the average difference was < 1 point (Mundt *et al.*, 2006).

The HAMD and MADRS have been compared in several publications, reaching different conclusions as to whether the MADRS is superior to the HAMD in detecting treatment effect (Maier and Philipp, 1985; Khan *et al.*, 2004; Carmody *et al.*, 2006).

## 4.3 The componential approach

The rating scales discussed in the previous sections provide only a global measure of disease severity. Therefore, Katz (1998) has argued that it is time for new methods to

test antidepressants. The lack of a well-defined measure of disease severity imposes an additional hurdle to the assessment of drug effect and its differentiation as novel targets are developed. In order to assess disease severity, it may be helpful to scrutinise the wording selected for the definition currently used by the WHO: "a common mental disorder that presents with depressed mood, loss of interest or pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, low energy, and poor concentration." Thus, it is clear that there are many symptomatic aspects underlying depression or a depressive state, besides mood alteration alone. Hence, strategies need to be developed to explore whether a depressive state exists only if mood is altered or if modulation of the other dimensions of the symptomatology can have equal impact on disease severity and hence on treatment response. If the assumption of interdependence between the dimensions of disease can be demonstrated, this would open a myriad of options for drug discovery. Instead of mood modulators, specific targets might be used to tackle other components of the disease.

Katz *et al.* (2004b) have developed a multivantaged behavioural method which uses 11 clinical scales to determine the severity of the different aspects of depression. They have identified three main components, anxiety-agitation-somatisation-sleep, depressed mood-motor retardation and hostility-interpersonal sensitivity. Because the measurement of all scales requires considerable time of both investigator and patient, a brief version of this approach was further developed which was shown to have similar characteristics (Katz *et al.*, 2004a). Interestingly, this paper also shows that antidepressants have specific effects on depression, and that they also differ markedly with respect to their time course. Despite these findings, there seems to be limited interest in the psychiatric research community to gather further insight into the dimensionality of disease and to pursue a strategy for the validation of clinical measures with focus on the assessment of efficacy.

#### 4.4 Imaging techniques

The lack of objective clinical measures of disease severity has created considerable expectation about the potential of brain imaging to elucidate pathophysiological features of depression, identify substrates for the placebo response and objectively quantify disease severity. In fact, imaging techniques have become more widespread in clinical research. Positron emission tomography (PET), which measures regional glucose metabolism with fluorodeoxyglucose (FDG), quantitative electroencephalography (QEEG) cordance and functional magnetic resonance imaging (fMRI) are the most important brain imaging techniques currently in use. Cordance is a measure derived from QEEG power, which has a strong association with cerebral perfusion and is measured by  $O^{15}$  PET.

FDG-PET and QEEG cordance have shown that the placebo effect is associated with specific and localised changes in brain functions (Leuchter *et al.*, 2002; Benedetti *et al.*, 2005; Haour, 2005). This is one of the few clinical investigations in which analysis is based on stratification of the findings (responder *versus* non-responder), rather than according to

treatment allocation (active *versus* placebo). FDG-PET scans of male patients with major depression responding to either fluoxetine or placebo after 6 weeks of study, show similar changes in metabolism (increases in prefrontal cortex and posterior cingulate and decreases in subgenual cingulate) (Mayberg *et al.*, 2002). Additional metabolic changes in the fluoxetine-responders are an increase in the pons region and decreases in caudate, insula and hippocampus regions. These additional regions provide an efferent input to the response-specific regions identified with both fluoxetine and placebo and could play a role in maintaining long-term clinical response and preventing relapse. On the other hand, no unique metabolic changes in placebo-responders are visible after 6 weeks.

The same investigation reveals that QEEG cordance of patients with major depression shows distinct brain changes in placebo-responders compared to medication (fluoxetine or venlafaxine) responders and non-responders. Placebo responders show a significant increase in prefrontal cordance already early in treatment which is not observed in medication responders, who show a decrease in prefrontal cordance. In placebo responders this effect becomes more marked in time, contrasting with the time course of medication responders, which is reduced over time (Leuchter *et al.*, 2002).

Even though these results may seem promising, all the research activities performed so far in this area have been exploratory and do not strategically address how such efforts can be incorporated into the development of new antidepressants. Validation of these imaging techniques is required to establish the value of the proposed measures as biomarkers for depression severity. To this purpose a correlation has to be established between the observed changes and overt symptomatology in a wide range of patients as well as in a control group under treatment with different drugs and placebo. Furthermore, discrimination must be made between controls and patients to determine the specificity of the findings and evaluate how sensitive these measures are to drug and disease-specific properties. This strategy would provide conclusive evidence about the potential for imaging measures to be used as surrogate endpoints in clinical trials. The availability of such a surrogate would however not eliminate the operational limitations of cost and maximum allowed exposure to radio-active materials. Alternative biomarkers or surrogate endpoints which provide an accurate, continuous measure of treatment effect may prove to be even more valuable in addressing essential research questions about drug efficacy.

## 5 ASPECTS OF CLINICAL TRIAL DESIGN IN DEPRESSION

In addition to the aspects concerning the sensitivity and specificity of the primary endpoint, trial design factors constitute another important source of confounders, leading to inaccurate estimates of drug effect. In general, clinical trials in depression are randomised, double-blind, placebo-controlled trials with a parallel group design. Usually, three treatment arms are included in the trial, consisting of placebo and two dose levels of a test drug, or placebo, a test drug and an active control. Between 100 and 250 patients are enrolled into each of these treatment arms with enrolment ratios between placebo

and drug varying from 1:1 to 1:2.5. The clinical endpoint is measured every 1, 2 or 3 weeks, with a higher measurement frequency at the start of the trial than at the end. Trial duration varies from 6 to 12 weeks, with the current general consensus that a 6-8 week duration is most appropriate (Montgomery, 2006).

Other important aspects are the inclusion and exclusion criteria. The most important criteria for clinical trials in major depressive disorder (MDD) are (1) a diagnosis of MDD, (2) a minimal HAMD score at entry from 18 to 20, which progressively increased in recent years, (3) no concomitant antidepressant or psychotropic medication and (4) no high risk of suicidal behaviour.

With respect to dosing regimens, various types of design can be applied. Frequently, titration designs are used, with an increase of dose level if a predefined effect is not achieved, and/or a decrease if side effects occur. Alternatively, a fixed dose design can be used, in which case no dose adjustments are allowed. The rationale for the selection of the dose level has not, however, deserved sufficient attention in early development. The concept of a maximum tolerated dose is still being used, irrespective of the pharmacological properties of the test drug. This situation is partly due to the lack of pre-clinical models of depression which can provide meaningful estimates of pharmacokinetic-pharmacodynamic relationships *in vivo*.

Another important aspect is the placebo response in depression. A commonly used design includes a run-in phase, which is aimed at reducing the placebo effect. The general set-up is as follows: investigators, but not patients are aware that for the first week of the trial only placebo is administered to all patients. If the HAMD (or another endpoint) decreases beyond a pre-specified threshold during this week, the patient is considered a placebo responder and subsequently excluded from the trial. A meta-analysis has shown that the outcome of studies with this type of run-in phase do not differ significantly from studies without a run-in phase, although the absolute effect size may be larger (Trivedi and Rush, 1994; Lee *et al.*, 2004). Another type of run-in phase is the double-blind placebo run-in phase, in which all patients are given placebo for a random time, unknown to both patient and investigator. The additional benefits of this clinical trial design element appear to be limited (Faries *et al.*, 2001). From the advancements on research about the placebo effect it has been established that placebo represents the effect of meaning (Moerman and Jonas, 2002). This definition raises questions about the influence of trial *context* as well as information provided in the informed consent forms on the patient. Clearly, a strategy is required to understand how this confounder can be better integrated and controlled, rather than excluded from the trial design.

Surprisingly, drug concentrations are not usually determined in antidepressant efficacy trials. Not only would such information facilitate the elucidation of the underlying exposure-response relationship, drug compliance could also be monitored. Various aspects of compliance can affect treatment outcome; the most important factors include poor execution and discontinuation of treatment (Vrijens and Urquhart, 2005). It has been shown that the use of diary cards is insufficient and unreliable (Stone *et al.*, 2002).

Electronic medication event monitoring (eMEM) systems have been developed to track patients' behaviour based on an electronic tag of the time and date in every occasion a container is opened (Urquhart, 1997). Given the social context of most depressed patients, one can easily conceive the implications of poor compliance on trial outcome. Close, blinded monitoring of patients' adherence and quality of execution are essential to obtain conclusive results about drug effect size.

## 5.1 Dropout

In all clinical trials a fraction of patients will not complete the pre-specified number of visits due to withdrawal or dropout. This may be due to lack of efficacy, the occurrence of side effects, or simply due to random events such as relocation or family circumstances. In clinical trials in depression, dropout rates vary between 1.6-61.3% (Machado *et al.*, 2006). In larger, well-controlled clinical trials however dropout rates between 20-30% are observed, depending on the duration of the trial. In statistical terms, three different mechanisms of dropout exist. The first mechanism is 'missingness completely at random' (MCAR). As this term indicates, this type of dropout mechanism is assumed to be independent of any variable, whether measured or not. The second mechanism is 'missingness at random' (MAR). Here, the probability of dropout depends on an observed variable, such as drug concentration or clinical endpoint. The third dropout mechanism is 'missingness not at random' (MNAR), where the probability of dropout is assumed to depend upon one or more unobserved variables. From a clinical perspective, one may assume that dropout for reasons related to lack of efficacy depend on the observed efficacy endpoint, and are thus MAR, or possibly related to unobserved efficacy endpoints (MNAR) (e.g., if a patient deteriorates after the last recorded measurement). Withdrawal due to side effects is most likely linked to exposure levels (MAR), and in absence of drug concentration data, as is the case in most large clinical trials in depression, may be considered a random event (MCAR).

The dropout mechanisms may cause biased estimates of drug effect, if not properly accounted for (section 6). These biases may lead to either an increased false positive rate, or an increased false negative rate (conservatism). Furthermore, expected dropout mechanisms are often not included in statistical power calculations, leading to underpowered studies.

## 5.2 Clinical trial simulation

Clinical trial simulation (CTS) is an important tool in evaluating the performance of different study designs without having to actually perform them. This methodology allows the formal incorporation of prior information about these factors into a quantitative framework which enables the creation of real life and extreme scenarios, providing an overview of the likelihood of possible outcomes for each situation. Therefore, it is not surprising that CTS is increasingly popular and many examples and reviews are available in literature (Girard, 2005; de Ridder, 2005; Kimko *et al.*, 2000). Two parts of the CTS approach

may be distinguished. Firstly, a drug-disease model needs to be formulated. This model describes how simulated patients will react to drug and placebo. The latter is important and highly variable in the case of depression. The second part of CTS is the trial execution model. Here, other aspects of real-life clinical trials are incorporated. Possible factors include patient compliance, dropout, protocol deviations and inclusion/exclusion criteria. Clearly, not all of these factors can be properly addressed in every CTS, since these largely depend on the drug-disease models that are used. Nevertheless, this approach offers an objective tool to assess the consequences of factors which are often evaluated by 'gut feeling', a standpoint that re-enforces preconceived beliefs and prevents innovation in clinical research. Unfortunately, the only two published examples of attempts to perform CTS to optimise clinical trial design in depression (Gruwez *et al.*, 2005, 2007) focused on limited trial design elements and did not take trial execution factors (e.g., dropout) into account.

## 6 STATISTICAL ANALYSIS OF DEPRESSION ENDPOINTS

In a conventional clinical study design, a primary endpoint has to be selected for the purposes of statistical inference or hypothesis testing. In efficacy trials, the primary endpoint is usually the clinical endpoint, e.g., a depression rating scale. It is important to note that in contrast to traditional statistic methods, model-based approaches exist in which multiple variables may contribute to the statistical inference. The next sections will discuss the various options that are currently available.

### 6.1 LOCF change from baseline

The analysis of variance (ANOVA) on change from baseline data at the end of a trial has been used historically and is still demanded by some regulatory agencies. In this method, last observation carried forward (LOCF) based imputation is used to accommodate patient dropout in the trial. Two important points of criticism may be made regarding this methodology. Firstly, a considerable amount of information that is available in the data is ignored since only the first and last measurements of each patient are taken into account. All other measurements in between are discarded. Secondly, LOCF imputation is subject to bias, depending on the mechanisms underlying dropout. If this mechanism is completely random, not dependent on any observed or non-observed variable (MCAR) *and* the dropout rate is the same in all groups, LOCF will be unbiased (Molenberghs *et al.*, 2004). If, however, the dropout is dependent on any variable, observed or unobserved (MAR or MNAR), or dropout rates are unequal, LOCF will be biased with the bias being determined by the general tendency of the data and the ratio of the dropout rates. In the case of a downward trend, as is the case in depression trials, the change from baseline may be underestimated.

## 6.2 Mixed model for repeated measures

In contrast to ANOVA, the mixed model for repeated measures (MMRM) (Mallinckrodt *et al.*, 2001b,a, 2004), a marginal linear mixed-effects model, enables the inclusion of all data collected during the trial to be used in the statistical analysis (Laird and Ware, 1982; Verbeke and Molenberghs, 2000). The observations for each individual are assumed to be drawn from a multivariate normal distribution, with different means for each time point and a correlation matrix which quantifies the correlation between the measurements at different times for each individual. This correlation matrix is assumed to be unstructured (it estimates the correlation between all measurements) and constant across all individuals and treatments. The treatment effect may be estimated as an average across all weeks, or only at the final measurement. Since this model takes into account all data, it leads to unbiased estimates under dropout mechanisms completely at random as well as dropout depending on observed data (MAR, i.e., the severity of depression) (Mallinckrodt *et al.*, 2003, 2004). Because of its robustness against these dropout mechanisms it is being used increasingly in the statistical analysis of clinical trials (Davis *et al.*, 2005; Kinon *et al.*, 2006; Thase *et al.*, 2006).

## 6.3 Hierarchical linear model

An alternative to the MMRM is the hierarchical linear model, or (single) random effects model (REM) (Laird and Ware, 1982; Verbeke and Molenberghs, 2000). Rather than the assumptions underlying the MMRM, the REM assumes a distinctive mean profile for each treatment, and a random subject-specific effect, which is usually chosen to be additive. This additive random effect is based on the assumption that each individual patient will experience scores which are on general above or below the mean profile. An advantage of this model is that it appears to fit individual data better because of the random effect. Additionally, its assumptions and parameterisation are easier to explain to clinicians and other non-statisticians, which may increase the acceptance of the model. Because of the characteristics of the normal distribution, the maximum likelihood estimates of the fixed-effects parameters of the marginal (MMRM) and hierarchical (REM) models are the same. However, the underlying assumptions allow different extensions of these models as will be investigated in the course of this thesis. Also, the hierarchical model has issues with respect to the approximation of the degrees of freedom (Molenberghs and Verbeke, 2004), which will be overcome in this thesis by using a fully Bayesian approach (see section 6.7).

## 6.4 Survival analysis

A completely different statistical approach to clinical trial data is the survival approach. This methodology was originally developed for areas where patient survival was the primary endpoint (such as oncology), hence the terminology. However, by using treatment response as endpoint rather than patient survival, these methods may be applied to clinical trials in depression. The strength of the survival approach is in determining the

time to a particular event, although adaptations exist which make it possible to parameterise a treatment effect in the proportion of non-responders (Chen *et al.*, 1999; Stassen *et al.*, 1993). The limited understanding of the sensitivity of the continuous scales currently used as endpoints in clinical trials, and the large variability in the time course of response, make the current definition of treatment response (a change of at least 50% from baseline) a natural dichotomisation. The most commonly used survival model is the non-parametric Cox proportional hazards model (Cox, 1972), which does not assume any shape on the underlying hazards, but proportionality between the hazards of the different patient groups throughout the observation window. Alternatively, parametric distributions may be used to describe the survival profiles. Despite a loss in flexibility, these models offer the option of extrapolation beyond the data and narrower confidence intervals of parameter estimates

### 6.5 Pharmacokinetic-pharmacodynamic modelling

Another approach to the analysis of clinical trials is that of pharmacokinetic-pharmacodynamic (PKPD) modelling. The basis of this approach is to elucidate the concentration-effect relationship for a particular class of compounds. This field is rapidly increasing in popularity and all major pharmaceutical companies currently employ these methods in pre-clinical research as well as during clinical drug development. One of the advantages of PKPD modelling is that extrapolation from animal models to human models becomes feasible once relevant targets and endpoints have been established (Yassen *et al.*, 2007). Also, because of the relation between drug exposure and response, it allows consistent evaluation of the influence of pharmacokinetic variability, in particular the effect of polymorphisms in metabolism (Goto *et al.*, 2007; Li *et al.*, 2007). Another important application is the ability to predict drug response associated with higher doses and different dosing regimens (Maas *et al.*, 2008). In conjunction with simulation techniques, it is possible to reduce attrition during drug development by showing how relevant factors, such as changes in drug formulations, modulate or alter drug response. Recently, PKPD modelling has moved beyond the characterisation of pharmacological effects or clinical response. More attention has been given to the incorporation of toxicity and safety endpoints, which allows dose optimisation taking into account both efficacy and safety.

In recent years, *mechanism based* PKPD models have gained in popularity, partly due to the increased meaningfulness of model parameterisation, and also because animal-to-human extrapolation is expected to be facilitated. Interestingly, in a recent thesis several mechanistic PKPD models have been applied to pre-clinical data in depression (Geldof *et al.*, 2007).

Unfortunately, several issues hinder the application of fully mechanistic PKPD models in depression. As mentioned before, the mechanism of action of antidepressants is at best partly known. Secondly, the large variability in the data prevents fitting complicated multi-parameter models to the data. Thirdly, since concentrations are not measured in larger clinical trials, linking drug exposure to drug effect is challenging. In literature,

one example is available of a semi-mechanistic KPD model, in which a dose-response relationship (rather than a concentration-effect relationship) is described (Gruwez *et al.*, 2007). Albeit a very interesting approach, a closer inspection reveals that the authors were unable to fit the data to the model (since one of the parameters is fit at the lower boundary). Because of the aforementioned reasons, PKPD models will not be used to fit data throughout this thesis. It is our endeavour to propose its use prospectively once potential sources of variability associated with disease and clinical trial design factors have been identified.

## 6.6 Functional data analysis

Functional data analysis (FDA) is an area that regards longitudinal data gathered from an individual as individual functions or curves. Whereas longitudinal data analysis is mostly concerned with the mean behaviour of populations, and the subsequent estimation of the differences between these groups, FDA is an exploratory discipline which is interested in the heterogeneity between individual curves, i.e., how patients differ from each other. Although FDA is not concerned with statistical inference, which explains its apparent absence from medical statistics, it may provide interesting insights into the variability between patients and may therefore lead to more appropriate statistical models.

FDA is normally applied to data which has been nearly continuously recorded over time. In some areas, this type of data is readily available. In depression however, it is clear that a visit frequency of once a week can hardly be considered continuous. Two approaches can be used to address this problem. Firstly, splines or other flexible curves may be applied to fit a continuous curve through the individual sets of data from each patient and subsequently utilise these results for FDA (Ramsay and Dalzell, 1991). Alternatively, a principal component analysis (PCA) may be applied on the original discrete data, where rather than performing dimension-reduction on measurements from multiple endpoints at a single time (as is usually the case), the dimensionality of multiple measurements (of a single endpoint) *over* time is reduced. The resulting principal components may then be smoothed (Rice and Silverman, 1991). In this thesis, an approach similar to the latter option will be proposed for the evaluation of patterns in the time course of response in depression.

## 6.7 Bayesian statistics

In addition to the classical maximum likelihood approach, which is typically applied to obtain parameter estimates in the aforementioned models, the Bayesian statistical framework is gaining in popularity in the field of statistical modelling in medicine. The key difference between classical and Bayesian statistics is the interpretation of probability.

In the classical framework, probability is considered a surrogate for frequency, hence the term 'frequentist' statistics. This means that a rigid statistician can only make probability statements about repeating events, such as coin tosses, but not about events that occur only once, such as the weather tomorrow or the outcome of a particular clinical

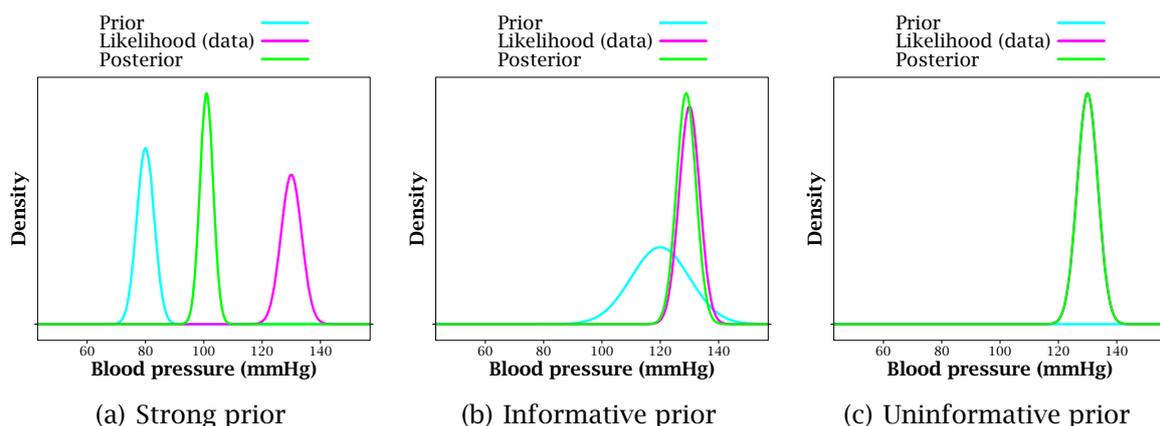
trial. This is somewhat circumvented by applying statistical methods many times, so that the frequency is in the use of the test rather than in the repetition of the experiment. An example of this is a confidence interval, which will contain the mean in 95% of the cases in which it is constructed (and not, as is commonly believed, contain the mean with 95% probability).

In a Bayesian context, probability is considered to be a quantification of 'degree of belief'. This means that Bayesian statisticians can make direct probability statements about any event, such as the probability that a treatment is superior to placebo.

The statistical reasoning behind this is that frequentist statistics relies on the likelihood, i.e., the probability of observing the data given a range of parameter values. Bayesian statistics however relies on the posterior distribution, which is the distribution of the parameters (i.e., parameter space) given the observed data. To obtain the posterior distribution, both the likelihood and a prior distribution are required. The prior distribution can include all relevant information that is not captured by the actual experiment. Since the use of this prior means that no two analyses will be the same, it is important to be explicit about the prior distribution that has been used in any Bayesian analysis.

An advantage of the resulting posterior distributions is that the interpretation of the probability is more direct. Indeed, 95% credible intervals (as they are referred to in a Bayesian context) do have the interpretation that there is a 95% probability of such an interval containing the true value. Figure 2 provides an illustration of the relationship between the prior distribution, the data (likelihood) and the posterior distribution.

The increase in the use of Bayesian statistics in recent years is explained by developments in computational sciences and enhanced computing capacity of computer proces-



**Figure 2.** Bayesian statistics: illustration of the prior-likelihood-posterior framework. The likelihood is the same in all situations and arose from two blood pressure measurements of an elderly woman. The posterior distribution resulting from three different prior distributions is shown. (a) Imagine a very strong prior, for example due to many blood pressure measurements in a twin sister. The posterior distribution will be heavily influenced by the prior distribution. (b) In this scenario, the population distribution of blood pressure is taken into account. Now, the prior is clearly less informative although some regression to the mean occurs. (c) When a non-informative flat prior is used, the likelihood and posterior distribution overlap. Example adapted from Spiegelhalter *et al.* (2004)

sors. For most Bayesian problems a closed-form analytical solution is not available. However, using Markov Chain Monte Carlo (MCMC) methods, it is possible to sample directly from the posterior distribution. Typically, at least 10,000 samples are required to characterise the posterior distribution. Increasing computer processor speeds have made this task feasible. It is important to emphasise that in the absence of an analytical solution, MCMC does yield unbiased results, as opposed to many of the maximum-likelihood based algorithms that are currently used. The development of WinBUGS (Lunn *et al.*, 2000) has removed many of the hurdles associated with the development of algorithms and paved the road for Bayesian statistics in many different areas.

## REFERENCES

- American Psychiatric Association (1952) *Diagnostic and statistical manual of mental disorders*, American Psychiatric Association, Washington, DC., 1st edition.
- American Psychiatric Association (1968) *Diagnostic and statistical manual of mental disorders*, American Psychiatric Association, Washington, DC., 2nd edition.
- American Psychiatric Association (1980) *Diagnostic and statistical manual of mental disorders*, American Psychiatric Association, Washington, DC., 3rd edition.
- American Psychiatric Association (1994) *Diagnostic and statistical manual of mental disorders*, American Psychiatric Association, Washington, DC., 4th edition.
- Angheliescu IG, Kohnen R, Szegedi A, Klement S, and Kieser M (2006) Comparison of hypericum extract WS (R) 5570 and paroxetine in ongoing treatment after recovery from an episode of moderate to severe depression: Results from a randomized multicenter study. *Pharmacopsychiatry* **39**:213–219.
- Artigas F, Nutt D, and Shelton R (2002) Mechanism of action of antidepressants. *Psychopharmacol Bull* **36 Suppl 2**:123–132.
- Atkinson AJ, Colburn WA, DeGruttola VG, DeMets DL, Downing GJ, Hoth DF, Oates JA, Peck CC, Schooley RT, Spilker BA, Woodcock J, and Zeger SL (2001) Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics* **69**:89–95.
- 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 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.
- Benedetti F, Mayberg HS, Wager TD, Stohler CS, and Zubieta JK (2005) Neurobiological mechanisms of the placebo effect. *J Neurosci* **25**:10390–10402.
- Bremner JD, Narayan M, Anderson ER, Staib LH, Miller HL, and Charney DS (2000) Hippocampal volume reduction in major depression. *Am J Psychiatry* **157**:115–117.
- Broadhurst A and Healy D (1996) Before and after imipramine, in *The Psychopharmacologist* (Healy D, ed.), pp. 111–134, Chapman & Hall, London.
- Carmody TJ, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D, Woo A, and Trivedi MH (2006) The Montgomery Asberg and the Hamilton ratings of depression: A comparison of measures. *Eur Neuropsychopharmacol* **16**:601–611.
- Castren E (2005) Opinion - is mood chemistry? *Nat Rev Neurosci* **6**:241–246.
- Chen MH, Ibrahim JG, and Sinha D (1999) A new Bayesian model for survival data with a surviving fraction. *J Am Stat Assoc* **94**:909–919.

- Cohn JB and Wilcox C (1985) A comparison of fluoxetine, imipramine, and placebo in patients with major depressive disorder. *J Clin Psychiatry* **46**:26-31.
- Cox DR (1972) Regression models and life-tables. *J R Stat Soc Ser B-Methodol* **34**:187-220.
- Czeh B, Fuchs E, and Simon M (2006) NK receptor antagonists under investigation for the treatment of affective disorders. *Expert Opin Investig Drugs* **15**:479-486.
- Davis LL, Bartolucci A, and Petty F (2005) Divalproex in the treatment of bipolar depression: A placebo-controlled study. *J Affect Disord* **85**:259-266.
- Delgado PL (2000) Depression: The case for a monoamine deficiency. *J Clin Psychiatry* **61**:7-11.
- Drew M and Hen R (2007) Adult hippocampal neurogenesis as target for the treatment of depression. *CNS Neurol Disord Drug Targets* **6**:205-218.
- Evans KR, Sills T, DeBrotta DJ, Gelwicks S, Engelhardt N, and Santor D (2004) An item response analysis of the Hamilton depression rating scale using shared data from two pharmaceutical companies. *J Psychiatr Res* **38**:275-284.
- Faries D, Herrera J, Rayamajhi J, DeBrotta D, Demitrack M, and Potter WZ (2000) The responsiveness of the Hamilton depression rating scale. *J Psychiatr Res* **34**:3-10.
- Faries DE, Heiligenstein JH, Tollefson GD, and Potter WZ (2001) The double-blind variable placebo lead-in period: Results from two antidepressant clinical trials. *J Clin Psychopharmacol* **21**:561-568.
- Fava M and Davidson KG (1996) Definition and epidemiology of treatment-resistant depression. *Psychiatr Clin North Am* **19**:179-&.
- Feighner JP (1999) Mechanism of action of antidepressant medications. *J Clin Psychiatry* **60**:4-13.
- Geldof M, Freijer J, van Beijsterveldt L, Vermote PCM, Meyens AA, and Danhof M (2007) Pharmacokinetic-pharmacodynamic modeling of the effect of fluvoxamine on p-chloroamphetamine-induced behavior. *Eur J Pharm Sci* **32**:200-208.
- Gelenberg AJ, Wojcik JD, Falk WE, Baldessarini RJ, Zeisel SH, Schoenfeld D, and Mok GS (1990) Tyrosine for depression - a double-blind trial. *J Affect Disord* **19**:125-132.
- Girard P (2005) Clinical trial simulation: A tool for understanding study failures and preventing them. *Basic Clin Pharmacol Toxicol* **96**:228-234.
- Goto S, Seo T, Murata T, Nakada N, Ueda N, Ishitsu T, and Nakagawa K (2007) Population estimation of the effects of cytochrome P4502C9 and 2C19 polymorphisms on phenobarbital clearance in Japanese. *Ther Drug Monit* **29**:118-121.
- Gruwez B, Dauphin A, and Tod M (2005) A mathematical model for paroxetine antidepressant effect time course and its interaction with pindolol. *J Pharmacokinetic Pharmacodyn* **32**:663-683.
- Gruwez B, Poirier MF, Dauphin A, Olie JP, and Tod M (2007) A kinetic-pharmacodynamic model for clinical trial simulation of antidepressant action: Application to clomipramine-lithium interaction. *Contemp Clin Trials* **28**:276-287.
- Guy W (1976) Clinical global impressions, in *ECDEU Assessment Manual for Psychopharmacology, revised* (Guy W, ed.), National Institute of Mental Health, Rockville, MD.
- Hamilton M (1960) A rating scale for depression. *J Neurol Neurosurg Psychiatry* **23**:56-62.
- Hamilton M (1967) Development of a rating scale for primary depressive illness. *Brit J Soc Clin Psychol* **6**:278-296.
- Haour F (2005) Mechanisms of the placebo effect and of conditioning. *Neuroimmunomodulation* **12**:195-200.
- Healy D (1997) *The antidepressant era*, Cambridge, Mass., [etc.] : Harvard University Press.
- Healy D (2006) Drug regulation - Did regulators fail over selective serotonin reuptake inhibitors?

- Br Med J* 333:92-95.
- Hindmarch I (2002) Beyond the monoamine hypothesis: mechanisms, molecules and methods. *Eur Psychiat* 17:294S-299S.
- Holsboer F, Lauer CJ, Schreiber W, and Krieg JC (1995) Altered hypothalamic-pituitary-adrenocortical regulation in healthy-subjects at high familial risk for affective-disorders. *Neuroendocrinology* 62:340-347.
- Jackson IMD (1998) The thyroid axis and depression. *Thyroid* 8:951-956.
- Kasper S, Anghelescu IG, Szegedi A, Dienel A, and Kieser M (2006) Superior efficacy of St John's wort extract WS (R) 5570 compared to placebo in patients with major depression: a randomized, double-blind, placebo-controlled, multi-center trial [isrctn77277298]. *BMC Med* 4:14.
- 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.
- 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, and Brown WA (2002a) Are placebo controls necessary to test new antidepressants and anxiolytics? *Int J Neuropsychopharmacol* 5:193-197.
- 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.
- Kimko HC, Reece SSB, Holford NHG, and Peck CC (2000) Prediction of the outcome of a phase 3 clinical trial of an antischizophrenic agent (quetiapine fumarate) by simulation with a population pharmacokinetic and pharmacodynamic model. *Clin Pharmacol Ther* 68:568-577.
- Kinon BJ, Lipkovich I, Edwards SB, Adams DH, Ascher-Svanum H, and Siris SG (2006) A 24-week randomized study of olanzapine versus ziprasidone in the treatment of schizophrenia or schizoaffective disorder in patients with prominent depressive symptoms. *J Clin Psychopharmacol* 26:157-162.
- Kline N (1959) Scales and checklists, in *Psychopharmacology: Problems in Evaluation* (Cole J and Gerard R, eds.), p. 475, National Academy of Sciences and National Research Council, Washington D.C.
- 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.
- Kraepelin E (1896) *Psychiatrie. Ein Lehrbuch für Studierende und Aerzte.*, J.A. Barth, Leipzig, 5th edition.
- Kramer MS, Winokur A, Kelsey J, Preskorn SH, Rothschild AJ, Snively 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.
- Kuhn R (1957) [treatment of depressive states with an iminodibenzyl derivative (G 22355)].

- Schweiz Med Wochenschr* **87**:1135-1140.
- Kuhn R (1958) The treatment of depressive states with G-22355 (imipramine hydrochloride). *Am J Psychiatry* **115**:459-464.
- Laird NM and Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* **38**:963-974.
- Lee S, Walker JR, Jakul L, and Sexton K (2004) Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation. *Depress Anxiety* **19**:10-19.
- Lenzer J (2006) Manufacturer admits increase in suicidal behaviour in patients taking paroxetine. *BMJ* **332**:1175.
- Leonard BE (2000) Evidence for a biochemical lesion in depression. *J Clin Psychiatry* **61**:12-17.
- 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.
- Leuner K, Kazanski V, Muller M, Essin K, Henke B, Gollasch M, Harteneck C, and Muller WE (2007) Hyperforin - a key constituent of St. John's wort specifically activates TRPC6 channels. *FASEB J* **21**:4101-4111.
- Li D, Lu W, Zhu JY, Gao J, Lou YQ, and Zhang GL (2007) Population pharmacokinetics of tacrolimus and CYP3A5, MDR1 and IL-10 polymorphisms in adult liver transplant patients. *J Clin Pharm Ther* **32**:505-515.
- Lieberman A (2006) Depression in parkinson's disease - a review. *Acta Neurol Scand* **113**:1-8.
- Linde K, Mulrow CD, Berner M, and Egger M (2005) St John's Wort for depression. *Cochrane Database Syst Rev* p. CD000448.
- Loomer H, Saunders J, and Kline N (1957) A clinical and pharmacodynamic evaluation of iproniazid as a psychic energizer. *Psychiatr Res Rep Am Psychiatr Assoc* **135**:129-141.
- Lunn DJ, Thomas A, Best N, and Spiegelhalter D (2000) WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat Comput* **10**:325-337.
- Maas H, Spruit M, Danhof M, and Della Pasqua O (2008) Relevance of absorption rate and lag time to the onset of action in migraine. *Clin Pharmacokinet* **47**:139-146.
- Machado M, Iskedjian M, Ruiz I, and Einarson TR (2006) Remission, dropouts, and adverse drug reaction rates in major depressive disorder: a meta-analysis of head-to-head trials. *Curr Med Res Opin* **22**:1825-1837.
- Maier W and Philipp M (1985) Comparative-analysis of observer depression scales. *Acta Psychiatr Scand* **72**:239-245.
- Mallinckrodt C, Clark S, Carroll R, and Molenbergh G (2003) Assessing response profiles from incomplete longitudinal clinical trial data under regulatory considerations. *J Biopharm Stat* **13**:179-190.
- Mallinckrodt C, Clark W, and David S (2001a) Accounting for dropout bias using mixed-effects models. *J Biopharm Stat* **11**:9-21.
- Mallinckrodt C, Kaiser C, Watkin J, Molenberghs G, and Carroll R (2004) The effect of correlation structure on treatment contrasts estimated from incomplete clinical trial data with likelihood-based repeated measures compared with last observation carried forward ANOVA. *Clin Trials* **1**:477-489.
- Mallinckrodt CH, Clark WS, and David SR (2001b) Type I error rates from mixed effects model repeated measures versus fixed effects ANOVA with missing values imputed via last observation carried forward. *Drug Inf J* **35**:1215-1225.
- Mayberg HS, Silva JA, Brannan SK, Tekell JL, Mahurin RK, McGinnis S, and Jerabek PA (2002) The functional neuroanatomy of the placebo effect. *Am J Psychiatry* **159**:728-737.

- Miller IW, Bishop S, Norman WH, and Maddever H (1985) The modified Hamilton rating-scale for depression - reliability and validity. *Psychiatry Res* **14**:131-142.
- Moerman DE and Jonas WB (2002) Deconstructing the placebo effect and finding the meaning response. *Ann Intern Med* **136**:471-476.
- Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, and Carroll RJ (2004) Analyzing incomplete longitudinal clinical trial data. *Biostatistics* **5**:445-464.
- Molenberghs G and Verbeke G (2004) Meaningful statistical model formulations for repeated measures. *Stat Sin* **14**:989-1020.
- 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.
- Moncrieff J, Wessely S, and Hardy R (2004) Active placebos versus antidepressants for depression. *Cochrane Database Syst Rev* p. CD003012.
- Montgomery SA (2006) Why do we need new and better antidepressants? *Int Clin Psychopharmacol* **21**:S1-S10.
- Montgomery SA and Asberg M (1979) A new depression scale designed to be sensitive to change. *Br J Psychiatry* **134**:382-389.
- Moore HK, Mundt JC, Modell JG, Rodrigues HE, DeBrotta DJ, Jefferson JJ, and Greist JH (2006) An examination of 26,168 Hamilton depression rating scale scores administered via interactive voice response across 17 randomized clinical trials. *J Clin Psychopharmacol* **26**:321-324.
- Mundt JC, Katzelnick DJ, Kennedy SH, Eisfeld BS, Bouffard BB, and Greist JH (2006) Validation of an IVRS version of the MADRS. *J Psychiatr Res* **40**:243-246.
- National Institute of Mental Health (2000) Depression, Technical report, Bethesda, MD, USA.
- Nielsen DM (2006) Corticotropin-releasing factor type-1 receptor antagonists: The next class of antidepressants? *Life Sci* **78**:909-919.
- O'Sullivan RL, Fava M, Agustin C, Baer L, and Rosenbaum JF (1997) Sensitivity of the six-item Hamilton depression rating scale. *Acta Psychiatr Scand* **95**:379-384.
- Pampallona S, Bollini P, Tibaldi G, Kupelnick B, and Munizza C (2004) Combined pharmacotherapy and psychological treatment for depression - a systematic review. *Arch Gen Psychiatry* **61**:714-719.
- Paykel ES (1985) The clinical interview for depression - development, reliability and validity. *J Affect Disord* **9**:85-96.
- Piccinelli M and Wilkinson G (1994) Outcome of depression in psychiatric settings. *Br J Psychiatry* **164**:297-304.
- Pierre OJ and Kasper S (2007) Efficacy of agomelatine, a MT1/MT2 receptor agonist with 5-HT2C antagonistic properties, in major depressive disorder. *Int J Neuropsychopharmacol*. pp. 1-13.
- Posternak MA and Zimmerman M (2005) Is there a delay in the antidepressant effect? a meta-analysis. *J Clin Psychiatry* **66**:148-158.
- Potts MK, Daniels M, Burnam MA, and Wells KB (1990) A structured interview version of the Hamilton depression rating-scale - evidence of reliability and versatility of administration. *J Psychiatr Res* **24**:335-350.
- Ramsay JO and Dalzell CJ (1991) Some tools for functional data-analysis. *J R Stat Soc Ser B-Methodol* **53**:539-572.
- Rees L (1960) Treatment of depression by drugs and other means. *Nature* **186**:114-120.
- Rice JA and Silverman BW (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. *J R Stat Soc Ser B-Methodol* **53**:233-243.
- de Ridder F (2005) Predicting the outcome of phase iii trials using phase ii data: A case study of

- clinical trial simulation in late stage drug development. *Basic Clin Pharmacol Toxicol* **96**:235–241.
- Rubin RT, Phillips JJ, Sadow TF, and Mccracken JT (1995) Adrenal-gland volume in major depression - increase during the depressive episode and decrease with successful treatment. *Arch Gen Psychiatry* **52**:213–218.
- Sapolsky RM (2000) Glucocorticoids and hippocampal atrophy in neuropsychiatric disorders. *Arch Gen Psychiatry* **57**:925–935.
- Sartorius N (2001) Concepts of depression: sporadic revolutions of continuous evolution. *Hum Psychopharmacol* **16**:S3–S6.
- Schildkraut J (1965) The catecholamine hypothesis of affective disorders: a review of supporting evidence. *Am J Psychiatry* **122**:509–522.
- Schoemaker C, Spijker J, and Poos M (2005) Hoe vaak komt depressie voor? [http://www.rivm.nl/vtv/object\\_document/o1275n17537.html](http://www.rivm.nl/vtv/object_document/o1275n17537.html).
- SFK (2006) Toename antidepressivagebruik stopt. *Pharm Weekbl* **141**:671–671.
- Sheline YI, Wang PW, Gado MH, Csernansky JG, and Vannier MW (1996) Hippocampal atrophy in recurrent major depression. *Proc Natl Acad Sci U S A* **93**:3908–3913.
- Sonino N, Fava GA, Belluardo P, Girelli ME, and Boscaro M (1993) Course of depression in Cushings-syndrome - response to treatment and comparison with graves-disease. *Horm Res* **39**:202–206.
- Spiegelhalter D, Abrams K, and Myles J (2004) *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Wiley, New York.
- Spijker J, De Graaf R, Bijl RV, Beekman ATF, Ormel J, and Nolen WA (2002) Duration of major depressive episodes in the general population: results from the netherlands mental health survey and incidence study (nemesis). *Br J Psychiatry* **181**:208–213.
- Stahl SM (1998) Basic psychopharmacology of antidepressants, part 1: Antidepressants have seven distinct mechanisms of action. *J Clin Psychiatry* **59**:5–14.
- Stassen HH and Angst J (1998) Delayed onset of action of antidepressants - fact or fiction? *CNS Drugs* **9**:177–184.
- Stassen HH, Delini-Stula A, and Angst J (1993) Time course of improvement under antidepressant treatment: a survival-analytical approach. *Eur Neuropsychopharmacol* **3**:127–135.
- Steffens D, Krishnan K, and Helms M (1997) Are SSRIs better than TCAs? comparison of SSRIs and TCAs: a meta-analysis. *Depress Anxiety* **6**:10–18.
- Stone AA, Shiffman S, Schwartz JE, Broderick JE, and Hufford MR (2002) Patient non-compliance with paper diaries. *Br Med J* **324**:1193–1194.
- Tedlow J, Fava M, Uebelacker L, Nierenberg AA, Alpert JE, and Rosenbaum J (1998) Outcome definitions and predictors in depression. *Psychother Psychosom* **67**:266–270.
- Thase ME, Macfadden W, Weisler RH, Chang W, Paulsson B, Khan A, and Calabrese JR (2006) Efficacy of quetiapine monotherapy in bipolar I and II depression - a double-blind, placebo-controlled study (the BOLDER II study). *J Clin Psychopharmacol* **26**:600–609.
- Trivedi MH and Rush J (1994) Does a placebo run-in or a placebo-treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacology* **11**:33–43.
- Urquhart J (1997) The electronic medication event monitor - lessons for pharmacotherapy. *Clin Pharmacokinet* **32**:345–356.
- Vallance AK (2007) A systematic review comparing the functional neuroanatomy of patients with depression who respond to placebo to those who recover spontaneously: Is there a biological basis for the placebo effect in depression? *J Affect Disord* **98**:177–185.

- Verbeke G and Molenberghs G (2000) *Linear Mixed Models for Longitudinal Data*, Springer-Verlag, New-York.
- Vrijens B and Urquhart J (2005) Patient adherence to prescribed antimicrobial drug dosing regimens. *J Antimicrob Chemother* **55**:616–627.
- Walsh BT, Seidman SN, Sysko R, and Gould M (2002) Placebo response in studies of major depression - Variable, substantial, and growing. *JAMA* **287**:1840–1847.
- Wang WF, Chang L, Minhas S, and Ralph DJ (2007) Selective serotonin reuptake inhibitors in the treatment of premature ejaculation. *Chin Med J (Engl)* **120**:1000–1006.
- Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu HG, Joyce PR, Karam EG, Lee CK, Lellouch J, Lepine JP, Newman SC, RubioStipek M, Wells JE, Wickramaratne PJ, Wittchen HU, and Yeh EK (1996) Cross-national epidemiology of major depression and bipolar disorder. *JAMA* **276**:293–299.
- Westenberg HGM and Sandner C (2006) Tolerability and safety of fluvoxamine and other antidepressants. *Int J Clin Pract* **60**:482–491.
- WHO (2007) WHO: Depression. [http://www.who.int/mental\\_health/management/depression/definition/en/](http://www.who.int/mental_health/management/depression/definition/en/).
- Williams J (2001) Standardizing the Hamilton depression rating scale: past, present, and future. *Eur Arch Psychiatry Clin Neurosci* **251 Suppl 2**:II6–12.
- Wong DT, Horng JS, Bymaster FP, Hauser KL, and Molloy BB (1974) Selective inhibitor of serotonin uptake - Lilly 110140, 3-(Para-Trifluoromethylphenoxy)-N-Methyl-3-Phenylpropylamine. *Life Sci* **15**:471–479.
- Yassen A, Ofsen E, Kan JM, Dahan A, and Danhof M (2007) Animal-to-human extrapolation of the pharmacokinetic and pharmacodynamic properties of buprenorphine. *Clin Pharmacokinet* **46**:433–447.
- Zobel AW, Nickel T, Kunzel HE, Ackl N, Sonntag A, Ising M, and Holsboer F (2000) Effects of the high-affinity corticotropin-releasing hormone receptor 1 antagonist R121919 in major depression: the first 20 patients treated. *J Psychiatr Res* **34**:171–181.
- Zung WWK (1965) A self-rating depression scale. *Arch Gen Psychiatry* **12**:63–70.