Chapter 1

Introduction

Depression and anxiety

Major depressive disorder (MDD) and anxiety disorders are highly prevalent mental disorders. The lifetime risk for MDD has been estimated at 23.2%, while that of any anxiety disorder has been estimated at 31.5% [1]. The core symptoms of MDD are depressed mood and anhedonia (loss of pleasure or diminished interest) [2]. The anxiety disorders form a heterogeneous group, which consists of generalized anxiety disorder (GAD), social anxiety disorder (SAD), obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD), panic disorder, agoraphobia, and specific phobia [2]. Their shared feature is the presence of anxiety states, varying from worry to obsessions to panic attacks.

Although there are clear differences among anxiety disorders and between anxiety disorders and depression, for instance in age of onset [1] and episode duration [3], they are also highly comorbid [4], which may be due to an underlying 'internalizing' liability [5]. Anxiety disorders, due to their earlier age of onset, are often a precursor to later depression [6], but depression can also precede anxiety disorders [7, 8, 9] and is often accompanied by clinically significant anxiety even in those without an anxiety disorder [10].

Treatment approaches for these disorders are also similar. First-line treatment strategies for both MDD and anxiety disorders include antidepressants and psychotherapy, primarily cognitive-behavioral therapy (CBT) [11, 12, 13, 14, 15, 16]. Although there are distinct CBT approaches for different disorders (e.g. exposure and response prevention for the treatment of OCD [17]), trans-diagnostic approaches have also been developed [18]. While antidepressants were initially developed and approved for the treatment of depression, they were later found to have similar efficacy for anxiety disorders [19, 20, 21]. They are now preferred over benzodiazepines, which were previously used frequently for anxiety but which have fallen out of favor due to their potential for abuse and dependence, particularly with long-term use [22]. Currently, slightly more than half of all antidepressant prescriptions are for MDD, while about a quarter are for anxiety disorders [23]. Antidepressant use has greatly increased over the last two decades, with 12% of the American population [24] and nearly 6% of the Dutch population [25] now receiving antidepressants yearly.

This thesis

Hundreds, if not thousands, of randomized controlled trials (RCTs) have been performed over the last half-century to demonstrate the efficacy of antidepressants and psychotherapy for the treatment of MDD and anxiety disorders. In spite of this wealth of evidence, however, essential questions remain unanswered.

In the first place, the evidence base is threatened by the presence of reporting and citation biases. As a consequence of these biases, the true efficacy and safety of these treatments

remains unclear, and this thesis therefore aims to re-examine the evidence base in order to clarify the impact of bias on the (apparent) efficacy and safety of these treatments. For methodological reasons, the emphasis will be on antidepressants, although many of these biases apply equally regardless of treatment modality.

Secondly, these treatments appear to be only modestly effective. While some patients respond very well to the first treatment that is tried (whether antidepressants, psychotherapy, or a combination), others require multiple treatment trials before finally, through a process of trial and error, chancing upon a treatment that is effective for them, and still others do not seem to respond well to any treatment. Further complicating the picture, some patients also show a good response to placebo, suggesting that for a subset of people suffering from MDD or anxiety disorders, some combination of the passage of time, supportive conversations, and instilling hope for improvement may be sufficient, without the need for any active (and potentially harmful) treatment.

Therefore, the second aim of this thesis is to use the existing evidence base to investigate clinical predictors of treatment response, in order to identify those patients who are likely to benefit from receiving (active) treatment prior to initiating treatment or as early as possible in the course of treatment.

Evidence-based psychiatry: the problem of bias

Although antidepressant use has increased over the past decade, these medications have also been embroiled in controversy. Much of this controversy has been due to allegations that pharmaceutical companies buried the results of trials that did not show that the drug was effective or that suggested that the drug was associated with safety concerns [19, 26, 27, 28, 29].

One of these trials, known as Study 329, has become particularly infamous. This trial examined the efficacy and safety of paroxetine, a selective serotonin reuptake inhibitor (SSRI), compared to imipramine, an older tricyclic antidepressant, and placebo in depressed adolescents. Originally conducted in the 1990s and published in 2001 [30], the results of this trial were used by SmithKline Beecham (SKB, now GlaxoSmithKline), the manufacturer of paroxetine, to promote the use of paroxetine in young people, with claims that the results showed the "REMARKABLE efficacy and safety" of paroxetine in the treatment of adolescent depression [28].

However, from the beginning there were concerns that the published article misrepresented the results of the trial [31, 32]. This was done by confusing the pre-specified primary outcome (which did not show significant efficacy of paroxetine) with a post-hoc secondary outcome that did yield a statistically significant result for paroxetine. Furthermore, serious adverse events, including suicide attempts, that occurred at an increased rate in the paroxetine group were dismissed as being unrelated to the drug. Re-analysis of

the trial by independent investigators concluded that neither imipramine nor paroxetine showed efficacy for adolescent depression in this trial and that both drugs were associated with an increased rate of adverse events, including suicidal behavior for paroxetine and cardiovascular problems for imipramine [33].

No other psychiatric trial has achieved the same notoriety as Study 329. However, bias is actually ubiquitous and has been demonstrated within psychiatry, medical science as a whole, and science in general [34]. Biased reporting can be motivated by commercial interests, as in the case of Study 329, but it can also occur in the absence of any commercial interests. Although much attention has (deservedly) been devoted to bias in antidepressant trials, for instance, there is every reason to believe that bias is just as common in psychotherapy trials [35]. In general, the scientific community appears to be prejudiced against null findings, so-called negative results that do not, for example, show that an intervention is effective.

Initial concern about bias mainly revolved around the problem of study publication bias, which was first identified more than fifty years ago [36] and which occurs when the likelihood of publication of a study depends upon the direction or significance of the results [37, 38]. However, negative results face additional obstacles to getting as much visibility as positive results than just study publication bias. These other biases include (but are not limited to) outcome reporting bias, spin, and citation bias.

Outcome reporting bias occurs when pre-specified outcomes are omitted from the published article, when new outcomes are added without being identified as new, or when the status of (non-significant) primary and (significant) secondary outcomes are switched. Although outcome reporting bias probably occurs regardless of study design, it is most easily identified in RCTs, because other studies (e.g. observational studies) often do not have a protocol with clearly defined primary and secondary outcomes. Outcome reporting bias has been detected in 31 – 62% of published RCTs [38, 39], indicating the scope of the problem.

Spin is defined as specific reporting strategies, whether intentional or unintentional, that could distort the interpretation of results. For example, while an article with spin does report non-significant results on the primary outcome, it nevertheless concludes that the intervention is effective on the basis of statistically significant results on secondary outcomes, in subgroups, or in pre-post comparisons. Among a sample of RCTs with non-significant results on the primary outcome, 58% contained spin in the conclusions section of the abstract and 50% contained spin in the conclusions section of the article itself [40]. Spin can affect clinicians' interpretations of a trial, leading them to judge interventions as being more beneficial [41].

Finally, citation bias refers to the tendency to preferentially cite positive studies, which has been demonstrated in various literatures [42, 43, 44, 45, 46]. Since studies are more likely to be discovered when they are cited by other studies, citation bias serves to high-

light positive results while negative results may go unnoticed.

Within the literature on depression and anxiety disorders, a landmark study was published in 2008, showing that negative trials of antidepressants for MDD were much less likely to be published than positive trials [19]. Furthermore, when these negative trials were published, they were often published as if positive. In this study, 51% of all trials were positive, but in sharp contrast, 94% of *published* trials appeared to be positive, and the effect size of antidepressants was overestimated by 32%. Study publication bias is also present in psychotherapy trials. Within a cohort of National Institutes of Health-funded psychotherapy trials, 24% of all initiated trials remained unpublished, and the effect size of these unpublished trials was markedly lower than that of published trials, clearly showing a bias against publishing unfavorable findings [35].

While most research on bias has focused on the efficacy of treatments, some research has also examined safety. Adverse events are rarely monitored, let alone reported, in psychotherapy trials [47, 48, 49]. Consequently, we know very little about the possible adverse effects of psychotherapy, although it is clear that psychotherapy, like all effective treatments, can have negative as well as positive effects [50, 51, 52]. Adverse events in drug trials, however, are usually extensively monitored because of regulatory requirements, but reporting on these events is limited, both in psychiatry [53] and in other medical fields [54, 55, 56]. Previous research on two antidepressants (sertraline and duloxetine) has suggested that harm outcomes are poorly reported and serious adverse events in particular are not always reported fully or accurately [57, 58]. In general, however, relatively little attention has been devoted to harm outcomes compared to efficacy outcomes.

Precision psychiatry: who benefits from treatment?

The research on bias in antidepressant trials has shown that these medications are less effective and also less safe than previously thought, particularly in young people. However, although antidepressant efficacy may, on average, be more modest than expected or desired, it is likely that some people do experience a robust and clinically meaningful response to antidepressants, while others experience very little, or no benefit at all. In general, around 15% more people respond to antidepressants than to placebo [59]. Consequently, there is great interest in predicting who benefits from treatment, to distinguish between patients who will recover even without active treatment, patients who will benefit specifically from the treatment, and patients who may need a different or more intensive treatment to recover.

Much effort has been devoted to examining biological and genetic markers that could be associated with antidepressant response. There has long been a particular interest in the serotonin-transporter-linked polymorphic region (5-HTTLPR), for instance, because SSRIs, the most commonly used antidepressants, inhibit the reuptake of serotonin by

blocking the serotonin transporter. Meta-analyses on this topic, however, have come to diverging conclusions on whether 5-HTTLPR is actually associated with antidepressant response [60, 61].

In general, candidate gene studies like these (in which a specific gene of interest is studied) have proven unlikely to replicate reliably [62, 63], due to the unfortunate combination of very small sample sizes (increasing the likelihood that any statistically significant finding is a false positive), analytical flexibility (providing the possibility of selecting the analysis or outcome that worked "best", that is, yielded a statistically significant result [64]), and reporting bias. For that reason, the field of genetics has largely abandoned candidate gene studies in favor of genome-wide association studies (GWAS), which have proven to be much more reliable [65].

In the past decade, GWAS have demonstrated that genetic effects are usually extremely small and very large sample sizes are required to detect these effects [62]. Obtaining such sample sizes within the context of a treatment trial is a major challenge. A recent meta-analysis of antidepressant pharmacogenetics trials, which included 2,256 participants, found only one genome-wide significant association; a polygenic risk score accounted for only about 1% of the variance in treatment outcomes [66]. Hence, it seems unlikely that genetic research will soon deliver a major contribution to the prediction of antidepressant response.

Other biological markers may have more potential, but so far decades of research have not resulted in any markers that are sufficiently predictive to be useful [67]. Biological markers based on electroencephalograms (EEG), functional magnetic resonance imaging (fMRI), or the like are also unlikely to be feasible in routine clinical practice. Markers that can be derived from blood or saliva tests may be more feasible, but still require additional steps that are seldom performed in clinical practice. Consequently, the requirements of feasibility dictate that routinely obtained and readily available clinical information should be used whenever possible, in preference over much less efficient alternatives like neuroimaging, unless this is shown to be much more accurate [68].

One obvious clinical characteristic is disorder severity. A 2008 meta-analysis suggested that initial severity of depressive symptoms was associated with antidepressant response, such that people with milder depression experienced little benefit from taking antidepressants compared to placebo [69]. Other research appeared to confirm this [70, 71, 72], and clinical guidelines were updated to reflect this finding and to recommend against the use of antidepressants as a first-line treatment for mild depression [73, 74]. More recently, however, two large studies did not replicate this association [75, 76], casting doubt upon this finding.

Although antidepressants are also commonly used for anxiety disorders, the evidence with regard to severity and antidepressant efficacy for anxiety is very limited, although some small studies have suggested that there is no association between initial severity and

antidepressant efficacy for anxiety disorders [77, 78]. Individual participant data meta-analyses have also found no evidence that initial severity of MDD moderates the efficacy of CBT compared to pill placebo [79] or antidepressants [80], in contrast to a widely-held belief that psychotherapy may be sufficient for mild depression, while antidepressants are necessary for severe depression.

Relatively few studies have examined whether clinical predictors could inform treatment selection [81]. Two studies have used data from the large STAR*D (Sequenced Treatment Alternatives to Relieve Depression) trial to predict who will (or will not) respond to antidepressant treatment using baseline clinical and demographic information [82, 83]. One of these studies [83] also found that the model appeared to be specific to treatment with citalopram and did not predict response to combined venlafaxine and bupropion, which suggests that it could be used for initial treatment selection. However, further validation and improvement would be necessary, as accuracy was still rather low (60%).

Studies in small samples have also suggested that simple clinical predictors could identify the optimal treatment for specific patients, when comparing antidepressants and CBT [84], or interpersonal psychotherapy and cognitive therapy [85]. The continued development of statistical learning techniques and the increased accessibility of large sample sizes through individual participant data meta-analysis are likely to offer opportunities to further improve these models.

However, at present, the possibilities for determining who will benefit from (which) treatment before the start of treatment are still limited. Another field of research, therefore, is aimed at investigating whether it is possible to determine whether a patient will benefit from treatment earlier in the course of treatment than is currently done. For instance, current guidelines usually recommend at least 4 and sometimes up to 8 weeks of treatment with an adequate dose of an antidepressant before considering switching antidepressants or augmenting treatment with other medications or psychotherapy [73, 74, 86]. However, many patients show a detectable, if modest, improvement within the first two weeks of treatment, and this early improvement has been robustly associated with attaining a full response or remission by the end of treatment [87, 88, 89, 90, 91, 92, 93].

While some of these studies have suggested that lack of early improvement is a sufficiently good predictor that a change in management is indicated for patients who do not show any improvement by two weeks [87], others have found that these patients still had a reasonably good chance of responding later in the study, which suggests that it would be premature to switch or augment treatment [94, 95]. Better predictive models, therefore, are desirable.

Recent research has suggested that symptoms are not interchangeable and the sum score on a depression or anxiety questionnaire may conceal important information [96]. For instance, symptoms such as sad mood or concentration problems appear to be more strongly associated with functional impairments than insomnia or changes in appetite

[97]. Improvement in specific symptoms is associated with a good response [98, 99, 100], but no study so far has controlled for the improvement in the sum score, so it is unclear whether examining individual symptoms can help to enhance predictive models.

Thesis outline

In this thesis, I aimed to bring the evidence base for treating depression and anxiety to light, particularly where it concerns antidepressants. In Part I, chapters 2 through 8, I study the impact of reporting and citation biases on the evidence base and also investigate whether the best available evidence (as synthesized in clinical guidelines) is actually put into practice. In Part II, chapters 9 through 11, I examine whether routine clinical information, specifically initial severity and early improvement in individual symptoms, can be used to predict who will benefit from antidepressants.

Part I: Bringing the evidence to light

Chapter 2 investigates reporting bias in clinical trials of antidepressants for the short-term treatment of anxiety disorders, focusing on the primary efficacy outcome. Using Food and Drug Administration (FDA) reviews, a complete cohort of trials was assembled and traced into the published literature to determine their fate. **Chapter 3** builds upon the results in chapter 2 by examining the oft-neglected flip side of the coin, namely safety. **Chapter 4** explores whether the practice of pooling trials for publication constitutes an additional type of reporting bias.

Chapters 5 and 6 examine spin (or positive focus) and citation bias. For this, I use the highly controversial literature on 5-HTTLPR, which has also been suggested to play a role in antidepressant response. I specifically look at a) gene-environment interactions in the development of depression and b) amygdala activation as an underlying mechanism for the development of depression. I examine whether spin and citation bias could play a role in the persistence of belief in spite of an unreliable evidence base.

The different reporting and citation biases are often examined separately, but in **Chapter 7** I examine the pernicious cumulative impact of study publication bias, selective outcome reporting bias, spin, and citation bias on the apparent efficacy of antidepressants and psychotherapy for depression.

This part concludes with **Chapter 8**, in which I study to what extent the evidence is actually put into practice. A prescription database is used to examine whether physicians adhere to the guidelines for antidepressant initiation in children and adolescents.

Part II: Who benefits from antidepressants?

Chapters 9 and 10 investigate the possible influence of initial severity on antidepressant efficacy for anxiety disorders, following the high-profile (but controversial) findings that antidepressants have minimal efficacy in mild depression. In **Chapter 9**, I look at whether the average baseline severity in a trial is associated with the antidepressant-placebo difference. As there are disadvantages to using trial averages, I study this question in more detail using individual participant data in **Chapter 10**.

Finally, in **Chapter 11**, I use early improvement to predict which depressed patients will benefit from antidepressants. In particular, I examine the predictive value of improvement in individual depressive symptoms, to determine whether examining individual symptoms can result in better predictive models than examining the total score alone.

Part I

Bringing the evidence to light