On the Effects of Serotonin Reuptake Inhibitors in Major Depression

Fredrik Hieronymus

2019

Department of Pharmacology,
Institute of Neuroscience and Physiology,
Sahlgrenska Academy, University of Gothenburg,
Gothenburg, Sweden

UNIVERSITY OF GOTHENBURG

Gothenburg 2019

On the Effects of Serotonin Reuptake Inhibitors in Major Depression

fredrik.hieronymus@neuro.gu.se
fredrik.hieronymus@gmail.com

ISBN 978-91-7833-312-7 (PRINT)
ISBN 978-91-7833-313-4 (PDF)
http://hdl.handle.net/2077/58234

Printed in Gothenburg, Sweden 2019
Printed by BrandFactory

Till Simone

On the Effects of Serotonin Reuptake Inhibitors in Major Depression

Fredrik Hieronymus

Department of Pharmacology,
Institute of Neuroscience and Physiology,
Sahlgrenska Academy, University of Gothenburg,
Gothenburg, Sweden

Abstract

This thesis focuses on the antidepressant effects of selective serotonin reuptake inhibitors (SSRIs) and how these are reflected by the Hamilton Depression Rating Scale (HDRS). To this end, we have assembled a large data set of placebo-controlled SSRI trials in major depression, and used this for a series of post-hoc patient-level analyses.

Thus, in a population of 8 262 patients treated with either of four SSRIs (citalopram, fluoxetine, paroxetine, or sertraline) or placebo, we have assessed (1) to what extent the various symptoms included in the HDRS separate between active treatment and placebo, and contrasted this to the sum-score of all HDRS items, which has been the conventional effect parameter, (2) whether the effects of SSRIs are dose-dependent, (3) whether side effects are necessary for SSRIs to outperform placebo, (4) if SSRIs increase or decrease suicidal ideation, and (5) whether only patients with high baseline depression severity respond to treatment with SSRIs.

We found that the influence of drug treatment on individual HDRS items differs vastly with regard to both the size and direction of effect. While *depressed mood* and other core symptoms of depression are consistently improved by SSRI treatment, HDRS items that may reflect typical SSRI side effects, such as e.g., *gastrointestinal symptoms* and *sexual symptoms*, respond, on average, negatively. The HDRS sum-score thus represents an aggregate of beneficial effects on core depression symptoms and detrimental effects on possible side-effect related items. Further, we suggest that the balance between these domains vary with time under treatment, with side-effects being relatively more influential

early in treatment, thereby obfuscating significant positive effects otherwise evident as early as after one week of treatment.

We also found evidence for a dose-response relationship, i.e., very low SSRI doses were more effective than placebo, but less effective than higher doses; this relation plateaued at the low to mid-range of currently recommended doses. We did not find any evidence in support of the hypothesis that side effects be an indispensable prerequisite for antidepressant efficacy, or that side effect severity moderates response. We could replicate previous studies showing SSRIs to decrease suicidal ideation in subjects ≥ 25 years of age, but could not detect a significant influence of SSRIs in either direction in young adults (18 ≤ age < 25). We found baseline symptom severity to be positively associated with SSRI efficacy when measured by the HDRS sum-score. This was however not the case for core depression symptoms where instead patients improved equally regardless of baseline severity. We suggest this to be partly due to non-core symptoms being absent in low-severity patients, thus leaving less room for improvement and more room for worsening on side-effect related items. Most of these observations were replicated in a population of 3575 patients from studies of the serotonin and noradrenaline reuptake inhibitor duloxetine.

We conclude i) that the sum-score of the HDRS rating scale is an insufficient and insensitive measure of antidepressant efficacy, ii) that the use of this outcome parameter has led to an underestimation of the true efficacy of SSRIs and SNRIs, particularly at the early phase of treatment and in subjects with relatively mild depression, iii) that normal doses of antidepressants are superior to low doses but not inferior to high doses, iv) that antidepressant effects are not, as has been suggested, secondary to side effects breaking the blind, and v) that the net effect of antidepressants on suicidality is beneficial, at least in subjects ≥ 25 years of age. In conjunction, the results rebut many of the claims that have been put forward by those questioning the usefulness of antidepressants.

Keywords: antidepressants, antidepressive agents, depression, major depressive disorder, selective serotonin reuptake inhibitor, SNRI, SSRI.

ISBN 978-91-7833-312-7 (PRINT)
ISBN 978-91-7833-313-4 (PDF)
http://hdl.handle.net/2077/58234

Sammanfattning på svenska

De selektiva serotoninåterupptagshämmarna (SSRI) utgör de oftast använda läkemedlen mot depression. Under senare år har deras antidepressiva effekt dock ifrågasatts, och det har även hävdats att de kan öka risken för självmord. Det forskningsprojekt som sammanfattas i denna avhandling har gått ut på att med nya infallsvinklar analysera utfallet av ett stort antal tidigare utförda placebo-kontrollerade studier vari SSRI-medel jämförts med placebo. I den första av avhandlingens sex artiklar undersöker vi SSRIs effekt på olika symptom, och finner härvid att nästan alla genomförda studier visar att dessa medel är tydligt bättre än placebo vad gäller att minska det centrala symptomet nedstämdhet. I den andra artikeln visar vi att låga doser av SSRI-medel är mer effektiva än placebo men sämre än högre doser. Att dessa låga doser ändå inkluderats i tidigare analyser har härmed medfört att man underskattat medlens positiva verkan. Vi ser också att SSRI, till skillnad från vad som ofta hävdas, utövar en liten men säkerställd förbättring av stämningsläget redan efter en veckas behandling. Den tredje artikeln undersöker om de har rätt som hävdar att enda skälet till att SSRI-medlen fungerar bättre än placebo i läkemedelsstudier är att biverkningarna får patienten att inse att han/hon ej lottats till placebo, vilket skulle kunna öka placebo-effekten hos dem som fått aktiv behandling. Vår observation att också de patienter som ej erfarit några biverkningar förbättras mer än de som erhållit placebo talar emot denna hypotes. I den fjärde artikeln undersöker vi hur SSRI påverkar självmordstankar hos patienter med depression. Vi finner härvid att medlen utövar en tydligt positiv effekt redan från första veckan hos dem över 25 års ålder, men att effekten hos yngre patienter är mer svårbedömd, delvis för att antalet yngre patienter i vår databas var förhållandevis litet. Den femte artikeln motiveras av att det ofta hävdas att det bara är patienter med mycket djup depression som eventuellt har nytta av SSRI-behandling. Vi finner dock att de med förhållandevis mild depression (men likväl av tillräcklig svårighetsgrad för att de skall ha inkluderats i läkemedelsstudier) vad gäller centrala symtom som t ex nedstämdhet förbättras i samma utsträckning som de med ett mer allvarligt tillstånd. I den sjätte och sista artikeln har vi granskat läkemedelsstudier avseende ett antidepressivt läkemedel ur gruppen SNRI-medel, dvs medel med delvis annorlunda verkan än SSRI-medlen, och kan härvid replikera de observationer vi gjort för SSRI-preparaten. Vi finner vidare att skillnaderna mellan SSRI och SNRI förefaller mindre än vad som tidigare hävdats.

List of papers

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. **Fredrik Hieronymus**, Johan F. Emilsson, Staffan Nilsson, Elias Eriksson. *Consistent superiority of selective serotonin reuptake inhibitors over placebo in reducing depressed mood in patients with major depression*. Mol. Psychiatry 2016;21:523-530.

II. **Fredrik Hieronymus**, Staffan Nilsson, Elias Eriksson. *A mega-analysis of fixed-dose trials reveals dose-dependency and a rapid onset of action for the antidepressant effect of three selective serotonin reuptake inhibitors*. Transl. Psychiatry. 2016;6:e834.

III. **Fredrik Hieronymus**, Alexander Lisinski, Staffan Nilsson, Elias Eriksson. *Efficacy of selective serotonin reuptake inhibitors in the absence of side effects: a mega-analysis of citalopram and paroxetine in adult depression.* Mol. Psychiatry. 2018;23:1731-1736.

IV. Jakob Näslund, **Fredrik Hieronymus**, Alexander Lisinski, Staffan Nilsson, Elias Eriksson. *Effects of selective serotonin reuptake inhibitors on rating-scale-assessed suicidality in adults with depression*. Br. J. Psychiatry 2018;212:148-154.

V. **Fredrik Hieronymus**, Alexander Lisinski, Staffan Nilsson, Elias Eriksson. *Impact of baseline severity on the effects of selective serotonin reuptake inhibitors in depression: an item-based patient-level post hoc analysis*. Submitted.

VI. Alexander Lisinski, **Fredrik Hieronymus**, Jakob Näslund, Staffan Nilsson, Elias Eriksson. *Item-based analysis of the effects of duloxetine in depression: a patient-level post hoc study*. Submitted.

Content

1 Background

1.1 Major Depressive Disorder

An episode of Major Depressive Disorder (MDD) consists of a period of at least two weeks in which the patient displays a pronounced reduction in overall mood and/or a marked loss of interest in, or pleasure from, usual activities (*anhedonia*).[1-3] To qualify for a diagnosis, at least three or four additional symptoms of depression (e.g., *weight change*, *sleep disturbance*, *fatigue*, *feelings of worthlessness*, *suicidal ideation*) need to be present. Furthermore, these symptoms must represent a change from the individual's usual state, they must cause significant distress or functional impairment, and they must not be better explained by another reason than depression, e.g., a somatic disease, substance abuse, or another psychiatric disease.

Lifetime prevalence of MDD varies considerably with sampling methodology and across populations, but has been estimated to be in the 10 to 20% range, with the 12-month prevalence being about half of that.[4-8] At least half of all patients who recover from their first depressive episode will have one or more additional episodes, and among patients who have had two episodes the recurrence rate is approximately 80%.[9] Depression is roughly twice as common in females as it is in men. Incidence peaks around middle-age, and being disabled, unemployed or of low income is associated with MDD, as is being previously or never married,[4-7] or having been exposed to traumatic events.[10]

Twin studies have estimated heritability to be around 40-50% and family studies indicate a two to threefold increase in lifetime risk of developing depression among first-degree relatives.[11,12] While progress in identifying genetic risk variants has been limited, a recent genome-wide association study including roughly half a million patients and controls identified 44 independent and significant loci associated with the development of MDD.[12]

MDD displays several psychiatric comorbidities and is hence strongly associated with substance abuse and dependence, various anxiety disorders (primarily generalized anxiety disorder), as well as some personality disorders.[6,7] MDD is also more common in patients suffering from chronic somatic diseases such as asthma, angina, arthritis, cancer, diabetes, or heart failure,[13-17] and it is also commonly seen post

myocardial infarction[18] and post stroke.[19] Patients with MDD have severely decreased quality of life and impaired psychosocial functioning,[13,20-23] are at a greatly increased risk of suicide,[21,24] and suffer a worse prognosis in several comorbid disorders.[19,25-27] A common disease, with a high recurrence rate, that is associated with considerable morbidity and mortality, MDD is thus one of the top contributors to the global burden of disease.[21,28]

1.1.1 Diagnosing depression

> "*[I]n 1953 the American Psychiatric Association held a 3-day 'Conference on the Development of a Research Program for the Evaluation of Psychiatric Therapies.' The goal of the conference was ambitious: to develop a 'comprehensive evaluation of therapies' and to establish 'sound criteria, methodology, or standards under which such validation might take place'. The conference ended with the frank admission that efficacy of treatment was impossible to judge because of the lack of standardized criteria for both diagnosis and treatment outcome.*"
>
> Mitchell Wilson, 1993

At the turn of the 19th century, the German psychiatrist Emil Kraepelin, by focussing on common patterns of symptoms over time, separated the unitary concept of psychosis into dementia praecox (schizophrenia) and manic depression (bipolar disorder), thereby establishing them as distinct disease entities.[29]

In contrast to this can be seen psychoanalytical theory, which dominated American psychiatry during the 1940s and 1950s, leading to a generally unfavourable view of psychiatric diagnoses.[30] A common argument at the time was that the precise symptomatic picture was secondary to understanding, in the words of Karl Menninger, '*how the observed maladjustment came about and what the meaning of this sudden eccentricity or desperate or aggressive outburst is.*'[31] In essence, effective psychiatric treatment thus consisted of understanding the meaning of a symptom and undoing its psychogenic cause (via psychotherapy), rather than manipulating a symptom directly (e.g., via

medication).[31] From this perspective, conventional diagnoses may even be seen as injurious to psychiatric patients as it shifts focus from the individual to the universal (i.e., the disease). Consequently, diagnostic agreement between psychiatrists at that time was poor.[32]

The ability to agree on which patients qualify for a diagnosis and which do not is an indispensable prerequisite for conducting meaningful quantitative research.[33] A landmark paper in this regard was published in 1972 by John P. Feighner and co-workers[34] in which they presented provisional operationalized diagnostic criteria for 14 psychiatric illnesses, which they argued had been '*sufficiently validated by precise clinical description, follow-up, and family studies to warrant their use in research as well as in clinical practice*'.

The core of the Feighner definition of depression was the presence of a minimum number of specific symptoms. This stands in stark contrast to the diagnosis of *depressive neurosis* in the then current DSM version (DSM-II), where the condition was defined as '*an excessive reaction of depression due to an internal conflict or to an identifiable event such as the loss of a love object or cherished possession*.'[35] The Feighner criteria for depression found their way, more or less unchanged, into the next revision of the DSM (DSM-III),[31,36] and thus our current conceptualization of depression as defined primarily by its symptomatology, or phenomenology,[37] was established. While there have been changes to the definition over the years, the similarity between the Feighner criteria and the current DSM-V diagnosis of Major Depressive Disorder is considerable.[2,34]

1.1.2 Problems with diagnosing depression

Though the switch from an etiological focus to a phenomenological one improved diagnostic reliability, depression is still a heterogeneous condition with comparatively low diagnostic reliability.[33] This was demonstrated in the field trials conducted during the development of the current iteration of the DSM (DSM-V) in which the MDD diagnosis was shown to have questionable interrater reliability ($\kappa = 0.28$).[38] Low diagnostic reliability is a problem as it introduces variance, thus making it more difficult to accurately identify e.g., risk factors and comorbidities, as well as potentially leading to erroneous inferences regarding e.g., the structure and natural course of the disorder.[33,39]

There are several possible reasons for why depression has comparatively low reliability. First, current consensus is that depression is best understood as a continuum with no sharp demarcation between well and unwell. And the same can probably be said for most of its constituent symptoms.[40,41] A diagnosis thus necessitates the imposition of a binary decision on an underlying continuous distribution, which invariably leads to borderline cases where judgements are likely to differ. Second, the many possible symptoms of depression make it possible for two patients to be diagnosed with MDD without having a single symptom in common.[42] Illustrating this, an analysis of a sample of 3703 depressed patients (the STAR*D population) found 1030 unique symptom profiles, with roughly half of them being endorsed by only one patient, and more than five out of six profiles being found in five or less subjects.[43] Third, MDD is strongly associated with several other psychiatric disorders and patients who qualify only for an MDD diagnosis is the exception rather than the norm.[44] As diagnoses are not hierarchical, and show considerable overlap in symptomatology,[2,45] the individual clinician has considerable leeway in deciding whether a patient who technically may qualify for several diagnoses is best described as suffering from all of them concurrently, or if his/her condition is better explained by one diagnosis being primary.[38]

A related concern is that the current diagnosis of depression is considered to have poor biological validity.[37,46-48] This has inspired initiatives aimed at reconceptualising psychiatric nosology into one based on biological understanding rather than phenomenology.[37] One such initiative is the National Institute of Mental Health-sponsored Research Domain Criteria framework, which proposes to move away from syndrome-based diagnoses and instead focuses on specific deficits in e.g., positive and negative valence systems, with possibly better-defined biological underpinnings.[49-52]

1.1.3 Measuring severity

If one were to follow the operational definition of depression, severe depression essentially means that the patient endorses more depressive symptoms, and/or that these symptoms are more severe, and/or that the patient's functional impairment is greater, as compared to what is the case for non-severe variants.[2,3] These criteria are unfortunately difficult to quantify and therefore not of much use for research purposes.[53] While

several proxies for severe depression, such as the presence of melancholia or hospitalization, have been suggested,[53,54] the most common way to address this issue has been to categorize patients based on some numerical measure of severity.[53,55]

To attain such a numerical measure, we require some transformation from clinical observations to numbers. This transformation can be as straightforward as letting an observer assign a reasonable numerical value to how severe he/she judges the condition of a patient to be, as in the Clinical Global Impression – Severity scale.[56] More commonly though, rating scales for MDD include a number of depression-related symptoms. Each symptom is graded on a scale by one or more observers, the scores for all included symptoms are summed together, and the sum-score is seen as representing the overall severity of the depressive episode.[57] By conducting repeated such evaluations it is then in principle possible to track the course of a depressive illness.

While there is a plethora of rating scales for depression in existence, only a handful have seen widespread use.[58,59] Pharmacological treatments of depression have primarily been evaluated using either the Hamilton Depression Rating Scale (HDRS), first published in 1960,[57] or the Montgomery-Åsberg Depression Rating Scale (MADRS), first published in 1979.[60] Psychological treatments, on the other hand, have mainly been assessed using the Beck Depression Inventory (BDI), first published in 1961.[61]

The HDRS, in its most common form, comprises seventeen items (HDRS-17), some rated on a 3-point scale and others on a 5-point scale. In order of appearance these items are: *depressed mood*, *feelings of guilt*, *suicidal ideation*, *initial insomnia*, *middle insomnia*, *late insomnia*, *work and interests*, *psychomotor retardation*, *psychomotor agitation*, *psychic anxiety*, *somatic anxiety¸ gastrointestinal symptoms*, *general somatic symptoms*, *sexual symptoms¸ hypochondriasis*, *loss of weight* and *insight*. The MADRS includes ten symptoms all of which are rated on a 7-point scale, the included items being: *apparent sadness*, *reported sadness*, *inner tension*, *reduced sleep*, *reduced appetite*, *concentration difficulties*, *lassitude*, *inability to feel*, *pessimistic thoughts* and *suicidal thoughts*. Finally, the BDI consists of 21 items which are all rated on a 4-point scale, the included items being: *sadness*, *pessimism*, *past failure*, *loss of pleasure*, *guilty feelings*, *punishment feelings, self-dislike*, *self-criticalness*, *suicidal thoughts or wishes*, *crying*, *agitation*, *loss of interest*, *indecisiveness*, *worthlessness*, *loss of energy*, *changes in sleep*

pattern, *irritability*, *changes in appetite*, *concentration difficulty*, *tiredness or fatigue* and *loss of interest in sex*.

While there is thus significant overlap between all three scales, there are also marked differences. The BDI stands out for being self-rated, although self-rated variants of both the MADRS and the HDRS do exist.[62,63] The BDI is also the scale with the largest emphasis on items related to affective cognition, whereas the HDRS includes the highest percentage of somatic items.[64] MADRS, in contrast to the BDI and the HDRS, was explicitly designed to be sensitive to change and hence included items that were frequently endorsed and which changed with treatment. Due to its symmetric structure and equal weighting of symptoms, the MADRS is generally considered to be a more balanced scale than the HDRS.[37,60] In this thesis we are primarily concerned with the HDRS for the simple reason that most clinical trials have used this scale as primary effect parameter.[65]

Notably, while many studies state that they have used the HDRS for outcome assessment, this is not as precise as it may seem. Max Hamilton himself published several versions of the HDRS[57,66] and other authors have produced versions that differ in which symptoms are included, as well as in how they are rated.[67,68] When Zitman and co-workers screened five major journals for one year, asking all authors of studies using the HDRS exactly which version they had used, they found that fewer than half of the publications referenced the version of the HDRS that had actually been used.[69]

1.1.4 Problems with the Hamilton Depression Rating Scale

The HDRS, though extensively used, has not escaped criticism.[58,64,65,70-73] Bagby and co-workers, when reviewing the scale in 2004, thus concluded '*[t]he HDRS is psychometrically and conceptually flawed. The breadth and severity of the problems militate against efforts to revise the current instrument.*'

One common criticism is that the HDRS is lacking in *content validity* and *face validity*, which essentially means that it is doubtful that the HDRS adequately and accurately measures all facets of MDD. Specifically, it has been pointed out that there is a considerable lack of

content overlap between the DSM diagnosis of depression and the HDRS assessment.[65] *Concentration difficulties* which are part of the DSM definition are not assessed by the HDRS, and while the HDRS includes an item on *guilt* it does not explicitly address *feelings of worthlessness* which is part of the DSM diagnosis. Moreover, in HDRS the anhedonia item is dominated by issues related to reduced ability to work rather than to a loss of interest and pleasure in various activities. The other way around, *psychic anxiety*, while common in depression and assessed by the HDRS, is not part of the DSM diagnosis and neither is *hypochondriasis* or *lack of insight*.[65]

As noted by Bagby and co-workers, given the dominant position of the HDRS one could almost as well criticize the DSM for not offering full coverage of HDRS-defined depression as the other way around.[65] However, due to the many different HDRS versions in circulation as many as 59 distinct symptoms have at some point been suggested to form part of an HDRS evaluation.[74] In the same vein, a review of seven commonly used depression rating scales found that they together included 52 different symptoms.[75] It is thus obvious that there are more aspects of depression, that could reasonably be measured, than those included in the HDRS-17. In analogy to the value of clearly defined diagnostic criteria, the substantial lack of content overlap between measures opens up the possibility that research outcomes may be conditional on which rating instrument is used.[75]

In contrast to *content validity* can be put *factorial validity*. While improving *content validity* essentially means including more symptoms, maximizing *factorial validity* aims to improve the correlation between the overt measure (i.e., the HDRS sum-score) and the latent construct (i.e., depression severity). In practice this often results in reduced *content validity* as only those symptoms that are strongly correlated to the latent construct are included.

To exemplify, if we wish to measure the latent construct ‘depression severity’ across multiple populations we would aim to exclude items that are strongly associated with other variables. For example, if certain HDRS symptoms are associated with age, gender, or a comorbid condition,[58,76] we run the risk of erroneously concluding that a particular patient group is more severely depressed when, in fact, some symptomatology may instead be attributed to other factors.[60,64,65] That rating scale sum-scores may reflect more than one underlying construct, i.e., that they can be *multidimensional*, is well-known, and Max

Hamilton himself placed considerable attention on the factor structure of the HDRS in his first publications on the scale.[57,66,77]

There are many good reasons to include many items in a rating scale: it increases *content validity* and *convergent validity,*[75] and it also tends to increase most measures of reliability even when inter-item correlations are weak.[78,79] This notwithstanding, a measure with adequate reliability and good content validity is still of little value as a change measure for depression if it is unclear whether it accurately represents depressive severity, as may be the case if there is low factorial validity.[60,65]

Further, even if all seventeen HDRS symptoms were equally relevant and informative with regard to depression severity, it does not necessarily follow that any particular treatment for depression must affect all symptoms equally.[58,64] Some antidepressants have insomnia as a side-effect[80] whereas others induce non-specific sedation by antagonizing histaminergic H_1-receptors.[81] Similarly, some more modern antidepressants may induce weight loss[82] and are associated with gastrointestinal complaints and sexual dysfunction.[54,80,83] All of these symptoms are rated on the HDRS; thus, a patient may in theory respond beneficially with regards to e.g., mood and cognition, score poorly on items reflective of side-effects, and have a null effect on the sum-score.[70] By using a scale including complaints that may be side-effects of treatment one may thus fail to distinguish between an effective treatment that causes side-effects and an ineffective one.

It is likewise not necessarily the case that all symptoms contribute equally with regards to disability.[84] A multivariate analysis of the impact of different depression symptoms on psychosocial functioning found that the proportion of total explained variance differed greatly between symptoms. While *hypersomnia* and *middle insomnia* accounted for less than 1% of total explained variance, *loss of interest*, *fatigue* and *concentration difficulties* stood for roughly 15% each, whereas *sad mood* contributed approximately 20%.[85] Supporting that not all symptoms are equal in this regard, it has been demonstrated that a substantial fraction of patients who score below the HDRS cut-off for remission do not consider themselves in remission,[86] and, conversely, that many patients who score above the HDRS cut-off for remission do not consider themselves depressed.[87]

One possible solution to these problem that could be applied retroactively would be to use factor scores from multidimensional scales

as outcome measures.[88] Another possibility would be to use unidimensional subscales, or single-item measures, derived from comprehensive rating instruments. Such subscales have seen some use in later years, especially the unidimensional HDRS-6 subscale[70,71] which comprises the items *depressed mood*, *guilt*, *work and interests*, *psychomotor retardation*, *psychic anxiety*, and *general somatic symptoms*.[58,70,71,89-91]

1.2 Pharmacological treatment of depression

> "*From the external appearance alone it is possible to tell that the mood improves with imipramine hydrochloride. The patients get up in the morning of their own accord, they speak louder and more rapidly, their facial expression becomes more vivacious. They commence some activity on their own, again seeking contact with other people, they begin to entertain themselves, take part in games, become more cheerful and are once again able to laugh ... despondency gives way to a desire to undertake something, despair gives place to renewed hope in the future. Instead of being concerned about imagined or real guilt in their past, they become occupied with plans concerning their own future.*"
>
> Roland Kuhn, 1958

The 1950s were transformational for psychiatry. In but a few years after the efficacy of chlorpromazine – the first antipsychotic – was established,[92] psychiatry received two pharmacological treatments for depression: iproniazid, the first monoamine oxidase inhibitor, and imipramine, the first tricyclic antidepressant (TCA).

That iproniazid, initially studied for its anti-tubercular efficacy, possessed psychoactive properties was first reported by Selikoff, Robitzek and Ornstein in 1952.[93] They noted that iproniazid, as compared to the structurally similar isoniazid, appeared a more potent stimulant of the central nervous system. Patients treated with iproniazid showed greater vitality, some to the point of wanting to leave the hospital.[94] At around the same time it was demonstrated experimentally

that iproniazid, but not isoniazid, inhibited monoamine oxidase, i.e., the enzyme primarily responsible for metabolizing monoamines such as serotonin, noradrenaline and dopamine.[95]

The observed psychostimulant effects, which were at the time primarily conceptualized as side effects, were by some discerning clinicians seen as a potential primary effect. In subsequent years, several reports concerning its potential as a treatment for depression were published.[96-98] Credit for the definitive assertion of iproniazid's antidepressant potential in depressed non-tuberculosis patients is usually afforded to Klein, Loomer and Saunders, who in 1957 reported that 70% of depressed patients had improved markedly (raised mood, weight gain, better interpersonal capacity, etc.) after treatment with iproniazid.[99] While the use of iproniazid was limited by side effects, compounds with similar monoamine oxidase inhibiting effects (e.g., tranylcypromine, phenelzine and isocarboxazid) were developed and are still being used.[94,100]

At around the same time, the Swiss psychiatrist Roland Kuhn was conducting a trial of a new compound designated G22355. Imipramine, as we now know it, was structurally related to chlorpromazine and was therefore investigated as a potential neuroleptic. While concluding that imipramine was not of much use for that purpose, Kuhn noticed a rapid and marked improvement in three patients diagnosed with depressive psychosis. He suggested it should be studied also in patients with endogenous depression, and thus the serendipitous discovery of the antidepressant properties of imipramine was made.[94,100,101] In 1961, Julius Axelrod and colleagues demonstrated that imipramine inhibited the reuptake of noradrenaline in peripheral tissue,[102] and in 1964 they showed this to be the case also in the brain.[103]

All this was paralleled by observations from Bernard Brodie and co-workers who in 1955 showed the Rauwolfia alkaloid reserpine to impact central stores of serotonin,[104] and by Holzbauer and Vogt who in 1956 showed that it did the same for central stores of noradrenaline.[105] Intriguingly, reserpine, which was at the time used in the treatment of high blood pressure, had several times been reported to induce depression in a subset of hypertensive patients.[106-109]

Together these three observations – that iproniazid and imipramine, who both increased noradrenergic activity via distinct mechanisms (decreased metabolism and decreased reuptake, respectively), could

alleviate depression, whereas reserpine that instead decreased noradrenergic activity, was able to induce it – formed the basis of the catecholamine hypothesis of affective disorders, popularized by Joseph Schildkraut in 1965.[110]

In 1968, Arvid Carlsson and colleagues demonstrated that imipramine also inhibits the reuptake of serotonin in the brain, and that this effect was evident at much lower doses than its effects on noradrenaline reuptake[111] which prompted Carlsson to remark that '*this action may be of importance for its antidepressive properties*'. One year later Lapin and Oxenkrug proposed a cohesive serotonergic theory of depression.[112]

TCAs, of which imipramine was the first, are 'dirty drugs', meaning that they interact with several targets.[113] Most TCAs block both the noradrenaline and the serotonin reuptake transporters, either directly or through their metabolites. Many also affect specific serotonergic and noradrenergic receptors, as well as have antihistaminergic and anticholinergic properties.[114] The TCAs in general have low therapeutic indices, produce significant side-effects, and are highly toxic in overdose.[115-117] Providing severely depressed patients with a substance potentially usable for self-poisoning was thus a concern,[118] and developing antidepressants with pure serotonin and/or noradrenaline reuptake inhibition became a priority.

The first selective serotonin reuptake inhibitor (SSRI) was zimelidine, which was developed by Berntsson, Carlsson and Corrodi and patented in 1972.[119] Reaching the market in 1982, it was shortly withdrawn after it was realized that it could occasionally trigger a Guillain-Barré-like syndrome.[120,121] By that time several other SSRIs were either already marketed or in late clinical development. The first SSRI to become a major commercial success was fluoxetine, which became available in 1986 and went on to be one of the world's best-selling drugs.[122]

Prescriptions for antidepressants rose dramatically during the early 1990s, primarily due to the SSRIs.[123,124] Fluoxetine was followed by several others (citalopram, paroxetine, sertraline, escitalopram), all of which achieved remarkable commercial success.[122] The reason why the SSRIs were rapidly adopted is not that they were more effective than prior generation antidepressants – in fact they may be less effective[125,126] – but because they were more tolerable and much safer in overdose.[127]

1.2.1 Methodological issues with antidepressant trials

Approximately half of all antidepressant trials conducted during the 1980s and 1990s have failed to show significant superiority of active treatment over placebo.[37,128-130] This has been deemed in contrast to earlier trials of antidepressants, many of which found significant differences despite sample sizes that were considerably smaller than the hundreds of patients per arm that are regularly included today, and which still often yield non-significant differences.[37,131] The high rate of failed and negative antidepressant trials has been a source of concern for the psychiatric community. Lower effect sizes necessitate the use of larger trials in order to attain statistical significance. Such trials are more expensive and time-consuming to conduct, and also likely to introduce additional sources of variability that further lowers statistical power.[37,84]

One possible explanation for the low rate of positive trials has been that the makeup of the patient population has changed. Prior to the introduction of effective antidepressants there existed a pent up need for effective treatment and it was thus logistically and ethically straightforward to recruit severely depressed patients to trials.[37,84] Now there are dozens of effective antidepressants readily available, and patients open to the possibility of receiving placebo, and which clinicians may consider to expose to the risk of not receiving treatment, are hence likely to be either less severely depressed or non-responders to previous treatments.[37,84] And while the definition of depression has not changed much since 1972,[34] the changes that have occurred have generally expanded the diagnostic boundaries – most recently through the removal of the bereavement exclusion.[33,37] It is thus likely that patients with milder, or more temporally variable, syndromes, and/or who have responded poorly to similar treatments, account for a larger fraction of the participants in current trials.[84] A related concern, though anecdotal, is that of professional malingerers who participate in multiple trials incentivized by the monetary rewards offered by some studies.[132]

Another explanation is that the studies have changed. Modern day trials are more complex than earlier ones, involving more frequent visits and mandatory evaluations, and they tend to carry on for longer durations. Thus more time and attention is invested in each patient and the increase in non-specific supportive contact may have increased placebo response rates.[84,132,133] A related concern, which has been argued could both inflate and underestimate antidepressant efficacy, is the strict inclusion

and exclusion criteria commonly used in antidepressant trials.[84,134] And, similar to the concern about professional symptomatic volunteers, it has been argued that pressure to include patients may incentivize clinicians to overrate the severity of some patients so as to make them eligible for inclusion.[135-139] If true, this would likely lead to decreased drug-placebo differences as all such overrated patients should display an apparent response to treatment as soon as the incentive to overrate them disappears, i.e., after they have qualified for inclusion.

1.2.2 Statistical issues with antidepressant trials

Two major statistical concerns when analysing data from clinical trials of antidepressants are how to deal with heterogeneous results across centres or trials, and how to handle missing data.[84,140] Heterogeneity is usually dealt with by formally assessing whether the results across, e.g., centres vary more than is expected by chance, and, if so, conducting a sensitivity analysis with the most extreme observations excluded.[84]

Missing data is more troublesome since there is no way to know how the observations that are missing would have turned out had they not been missing.[140] While missing data can arise through multiple mechanisms, the by far most common cause is that subjects drop out of the trial, usually due either to adverse events or lack of efficacy of the randomized treatment. Since antidepressants have a gradual onset of efficacy, a large fraction of early drop-outs is a major issue that can severely bias results.[84]

In statistical parlance three terms are commonly used to describe missing data: MCAR (missing completely at random), MAR (missing at random), and MNAR (missing not at random).[141] MCAR means that there are no systematic differences between missing and observed data. The missing data is truly a random subset of all observations and estimates are hence not biased by their omission. For MAR data there are systematic differences between missing and observed data, but those differences are conditional on some observed variable. If, for example, low baseline quality of life contributes to both dropout and poor treatment outcome, this would bias an unadjusted model. However, if the model controls for baseline quality of life then estimates will be unbiased. If data is MNAR this means that there are systematic differences between missing and observed data, and that this difference

is not conditional on any observed variable. Consequently, we have no way to adjust for the missing data and there will be systematic deviations between observed and unobserved data. With the exception of obvious cases of MCAR data, it is generally impossible to distinguish between data that is MAR and data that is MNAR.[141]

Missing data has historically been handled in one of two ways. It has either been ignored, in a so-called observed cases (OC) analysis, or a single imputation approach, commonly the last observation carried forward (LOCF) procedure, has been used. The population generated by this procedure, when conducted on all patients who have been randomized to receive treatment, is called an intention-to-treat (ITT) population. Neither of these methods is valid if data is MNAR or MAR, and LOCF is not necessarily valid if data is MCAR since it assumes that no change would have occurred following dropout. Nevertheless, LOCF has been commonly used and is generally seen as a conservative estimate of treatment effects, which, however, is not necessarily the case.[142,143]

Recently, linear mixed-models and multiple imputation methods have seen more widespread use.[144] The advantage of these over OC and LOCF being that they produce unbiased estimates when missing data is MCAR or MAR.[141-143] When data is MNAR it is impossible to obtain unbiased estimates and the appropriate way to handle this possibility is to conduct a number of sensitivity analyses using, e.g., pattern mixture models or selection models.[142]

1.2.3 Statistical and clinical significance

The magnitude of effect needed to qualify as meaningful is usually referred to as clinical significance, or clinical relevance.[145] Statistical significance, on the other hand, quantifies how likely it is that an observed difference is due to chance. Statistical significance, in contrast to clinical significance, is affected not only by the size of the treatment effect but also by the sample size used in the particular investigation.[145] Provided that there is an underlying true effect, no matter how minute, the probability of attaining statistical significance approaches one-hundred percent as the sample size increases. For this reason, statistical significance by itself is not necessarily informative.

Unfortunately, clinical significance can be equally uninformative since there is no consensus on what separates clinically significant from clinically insignificant, and the same result can hence be used to argue both for and against the usefulness of any particular treatment.[146-148] An example of this is constituted by two antidepressant meta-analyses, published within a month of each other by Turner and colleagues in the New England Journal of Medicine[130] and by Kirsch and co-workers in PLOS Medicine.[149] The former, which found a publication bias-corrected effect size of 0.31, concluded that all included antidepressants were superior to placebo. The latter, which found a highly similar publication bias-corrected effect size of 0.32, instead concluded that: '*there seems little evidence to support the prescription of antidepressant medication to any but the most severely depressed patients*'.[147] Additionally, since non-response and treatment-resistance to antidepressants are common occurrences,[150,151] a metric such as 'average additional improvement for all patients' (i.e., effect size), likely underestimates the benefit afforded to those patients who do respond.

1.3 The antidepressant controversy

Due to the lack of useful therapeutic tools in the treatment of depression, the first antidepressants were rapidly adopted and have proven to be extraordinarily useful;[94,118,122-124] a 1975 review on imipramine concluding that '*[t]he benefit of this drug in patients with endogenous depression who have not become institutionalized is indisputable … further drug-placebo trials in this condition are not justified.*'[131] Nevertheless, they have since their introduction also been the target of criticism.[54,152] During the 1960s and 1970s, in addition to the tensions between psychodynamic psychotherapy and biological psychiatry, antidepressants were targeted by the antipsychiatry movement,[153] as well as by the Church of Scientology whose founder, L. Ron Hubbard, blamed psychiatry for everything from being the primary cause of crime[154] to creating the holocaust.[155]

In later years, the criticism against antidepressants has become both more prevalent and more public than before.[156-161] What was previously essentially an intra-professional debate on, e.g., the merits of antidepressants in milder forms of depression,[84,127] has thus transformed into cover stories categorically stating that antidepressants '*don't work*'.[160] That this stance has gained popularity also outside a limited

group of outspoken critics is illustrated by a book review where a former editor in chief of the New England Journal of Medicine entertains the prospect that antidepressants are '*[u]seless ... or worse than useless.*'

Among the more vocal critics of antidepressants can be mentioned Harvard psychology professor Irving Kirsch, whose stance is that the SSRIs do not display beneficial effects, and that the effects of antidepressants in clinical trials are artefactual and caused by inadequate blinding.[162,163] In the same vein, the British psychiatrist Joanna Moncrieff argues that antidepressants do not have any specific effects, and that any apparent effect in trials may be explained by the rating scales used being sensitive to non-specific effects such as sedation.[72,164,165] While David Healy, professor of psychiatry at Bangor University, has claimed that the SSRIs commonly induce suicide and violent behaviour,[166,167] the American journalist Robert Whitaker claims that antidepressants, and other psychiatric drugs, lead to brain damage and malign long-term outcomes and thus are the cause of an iatrogenic epidemic of mental illness.[168] And recently this debate has witnessed the entry of the former director of the Nordic Cochrane Center, Peter Gøtzsche, who argues for all of the above, with the addition that antidepressants also induce dependence.[169]

The controversy has not only affected the public perception of these drugs but has also influenced regulatory authorities, such as when a representative of The National Institute for Health and Care Excellence in Britain stated on CBS 60 Minutes that antidepressants '*probably weren't worth having*' for mild to moderate depression.[156] Or when a Special Rapporteur to the U.N. Human Rights Council, in his report on '*the right of everyone to the enjoyment of the highest attainable standard of physical and mental health*' concluded that '*[t]he benefit experienced with antidepressants, specifically for mild and moderate depression, can be attributed to a placebo effect*.'[170]

1.3.1 Antidepressants and suicide

Early epidemiological analyses from the United States concluded that while the rate of suicide was relatively stable in the decade following the introduction of the first pharmacological antidepressants, there was a marked increase in suicide attempts.[171] This finding prompted some

authors to speculate that perhaps pharmacological antidepressants were less effective at preventing suicide than electroconvulsive therapy.[172]

A later theory posited that these observations, rather than being a consequence of poorer anti-suicidal efficacy, may in fact be due to a suicide-promoting effect. Specifically, it was suggested that antidepressants with predominantly noradrenergic effects could reduce psychomotor retardation and inhibition prior to there being any marked improvement in mood or suicidality,[173] hence potentially affording severely depressed and inhibited patients the energy to go through with a suicide attempt that might otherwise be lacking.

Shortly after the introduction of the SSRI fluoxetine, several case reports detailing emergence of intense suicidal ideation after initiation of therapy were published. What was described was distinct from the previous disinhibition theory: an intense '*somatic-emotional state*' of profound anxiety and/or restless agitation in combination with an inability to sit still (akathisia) and severe suicidality.[174-179] Scientology, through its antipsychiatry organization, the Citizens Commission on Human Rights, took heed of these reports and mounted major and partially successful campaigns against fluoxetine through television appearances and newspaper ads.[154,180]

The case reports on possible fluoxetine-induced suicidality prompted Eli Lilly to conduct a meta-analysis of their fluoxetine trials. The report, which was published in the BMJ in 1991, concluded that there were no statistical differences in the rates of suicidal acts between patients treated with fluoxetine, TCA comparators, or placebo, but that suicidal ideation instead improved significantly more often with fluoxetine or TCAs than it did with placebo ($p < .001$).[181,182]

Since then this issue has been the subject of several meta-analyses. Summarizing these is not entirely unproblematic due to the lack of standardized methodology to investigate suicidality. Different publications thus report various composite outcomes consisting of suicidal ideation and/or suicidal behaviour and/or suicide attempts and/or completed suicides.[183] While individual studies have suggested that certain antidepressants, such as maprotiline, may be associated with an increase in completed suicides,[184] suicide attempts,[184] and suicidal ideation,[185] authors reviewing the literature have generally not been able to detect any differences in the rates of completed suicides.[186-191] Regarding suicide attempts and/or suicidal behaviour, some studies have

found an increase[187,192] while others have not.[188,190,193-195] Effects on suicidal ideation, while seldom analysed separately, have tended to be equivocal[190] or positive.[182,193,196] Finally, studies looking at rating scale-assessed suicidality, which depending on the particular scale may be seen as an amalgamation of suicidal ideation, behaviour and attempts,[197] have generally found beneficial effects of antidepressants.[182,193-195,198,199]

A highly influential publication on this matter is a report from the U.S. Food and Drug Administration which was published in the BMJ in 2009,[200] in which Stone and co-workers present safety analyses comprising roughly 100000 adult patients who had participated in 372 trials of antidepressants, regardless of indication. Prompted by previous internal analyses of paediatric antidepressant trials, which had found an increase in suicidality (but not completed suicides, of which there were none),[192] they conducted age-stratified analyses which found a significant increase in suicidal behaviour, but not ideation, in young adults ($18 \leq \text{age} < 25$), a neutral effect on suicidal behaviour and a positive effect on suicidal ideation in participants aged 25 to 64, and a positive effect on suicidal behaviour and a similar tendency for suicidal ideation in those above the age of 65. Corresponding age-dependent effects have been reported also from observational studies.[201,202]

Based on their review of paediatric trials,[192] the U.S. Food and Drug Administration in 2004 issued a black-box warning about the possibility of a link between antidepressant use and suicidality in children and adolescents. In 2007 the black-box warning was expanded to also include young adults ($18 \leq \text{age} < 25$), and regulatory agencies around the world followed suit. As depression is a major risk factor for suicide also in children, adolescents, and young adults, but also since such a warning may have spill-over effects to other age-groups, this decision was highly controversial.[203-211] While there were reports of an increase in adolescent suicides and suicidal behaviour following the black-box warning,[212-214] similar trends were not seen in all countries.[215,216]

A reasonably consistent finding across countries, however, is that of an inverse relation between antidepressant prescriptions and suicides.[123,217-219] Similarly, Gibbons and associates, in a sample of roughly a quarter of a million American veterans, found the rate of suicide attempts to be lower in antidepressant-treated patients than in those who received no treatment, and that the rate of suicide attempts was higher prior to treatment than after treatment initiation.[220] Paralleling this, Isacsson and co-workers have repeatedly shown that

suicide subjects are unlikely to test positive for antidepressants also in the presence of major depression,[221,222] and Rutz and colleagues found that a physician's education program administered on the Swedish island of Gotland and aiming to improve detection and treatment of depression showed time-related drops in suicide rates.[223,224]

Gibbons and colleagues also demonstrated that, on the county level, SSRI prescriptions were inversely related to suicidality whereas TCA prescriptions were positively related,[225] which could partly be explained by SSRIs being less toxic in overdose.[118] Similarly, Tiihonen and co-workers, in a cohort of 15390 patients hospitalized due to a suicide attempt, found that among patients who had ever used an antidepressant, current use was associated with an increase in suicide attempts, but a decrease in completed suicides.[226]

Due to the low rate of suicides in clinical trials of antidepressants it has been estimated that 1.9 million participants would have to be enrolled for a study to have adequate power to detect a 20% increase or decrease in completed suicides.[188] It is thus unlikely that this matter will be definitively resolved by any future randomized controlled study; hence thoroughly analysing data which have already been collected for indications of whether these drugs may be harmful in a subset of patients should be a priority.

2 Aims

The overall aim of this thesis is to investigate the effects of SSRIs when used as a treatment for MDD. Specifically, we look to assess what influence the common use of the HDRS sum-score for evaluating treatment outcomes may have had on apparent SSRI efficacy. A common feature of the six papers is that we have assessed multiple HDRS derived efficacy measures, including individual items, and contrasted the results obtained from these with those arrived at when using the HDRS sum-score. In addition, item-based analyses were used to assess possible symptom aggravation.

In paper I we aimed to assess whether the use of the HDRS sum-score as primary outcome measure has contributed to why roughly half of all trials have failed to show a significant separation between SSRI and placebo.

In paper II we wished to investigate whether the beneficial effects of SSRIs are dose-dependent, and to what extent including suboptimal doses in post-hoc analyses makes the drugs appear less effective than they are.

In paper III we aimed to assess whether side-effects are necessary for SSRIs to show superiority over placebo, i.e. to test the validity of the side-effects-breaking-the-blind theory.

In paper IV we attempt to assess the impact of acute SSRI treatment on rating scale-assessed suicidal ideation in young adults ($18 \leq \text{age} < 25$) and adults ($\text{age} \geq 25$), respectively.

In paper V we investigate whether baseline depression severity impacts SSRI efficacy. Specifically, we aimed to investigate whether patients with comparatively mild depression benefit from SSRI treatment.

In paper VI we aimed to see whether the results we had for the SSRIs would extend to an SNRI. Also, we wanted to compare the effect profile of the SNRI to that of the SSRIs after adjusting for overall antidepressant efficacy.

3 Papers

Papers I to V are based on patient-level data from a population of phase II to phase IV trials conducted by the pharmaceutical industry during the clinical development of three SSRIs: citalopram (H/S Lundbeck, Valby, Denmark), paroxetine (GlaxoSmithKline, Brentford, UK) and sertraline (Pfizer, New York, NY, USA). Paper VI is similarly based on trials conducted by Eli Lilly (Indianapolis, IN, USA) during the clinical development of duloxetine.

The data base used for most analyses in papers I-V consists of 8262 patients from 28 SSRI trials which have used the HDRS for rating purposes. Participants were treated with either citalopram (n=744), fluoxetine (n=754), paroxetine (n=2981), sertraline (n=1202) or placebo (n=2581). The data used for paper VI comes from 15 studies on duloxetine comprising 4828 patients. Of these 2709 were treated with duloxetine, 202 with escitalopram, 60 with fluoxetine, 290 with paroxetine and 1559 with placebo.

In paper I we conducted analyses of individual studies and therefore excluded trials with small sample sizes (n<50 for each arm). The complete sample for this paper thus consists of 18 trials and 6669 eligible subjects. In paper II we specifically looked at dose-response in fixed-dose trials (n=11) including in total 2859 subjects. For paper III we needed data on the timing and nature of adverse events which we had not requested in our initial research proposals. Pfizer was unable to provide us with these data and we were hence unable to include studies of sertraline. Lundbeck provided additional data in an offline format, thus allowing us to use the same data sets as for all other analyses of citalopram. GlaxoSmithKline, on the other hand, had implemented new procedures for data sharing and provided remote desktop access to the requested data. While the included studies are the same, there may exist minor discrepancies between the online and the offline data sets. Papers IV and V use the full SSRI data set, and paper VI uses the complete duloxetine set.

3.1 Paper I

3.1.1 Background

The high rate of failed antidepressant trials has been a major concern for the psychiatric community, in part because it makes new drug development more expensive and hence imperils future drug discovery, but also because it has been used as an argument to suggest that these drugs are not clinically relevant (see 1.2.1).[162,163,227,228]

Prompted by the problems associated with HDRS sum-scores (see 1.1.4) we aimed to investigate what influence the use of this measure as primary effect parameter may have had on the poor outcome of many SSRIs trials.[229] As we had access to item-level data for individual patients we could compare the results obtained when using *HDRS-17-sum* as outcome measure, to those obtained when alternative HDRS-derived unidimensional subscales and single-item measures are used. We particularly emphasized *depressed mood* since this measure has good face validity (one of two cardinal symptoms), was present in almost all patients (highest baseline rated severity) and should hence not be subject to large floor effects, and has previously been used as a secondary efficacy measure by regulatory authorities.[230]

3.1.2 Results

With regards to the results from individual studies, 56% of all comparisons (18 of 32) failed to separate between active drug and placebo when *HDRS-17-sum* was used as outcome measure, as compared to 9% (3 of 32) when *depressed mood* was used for the same purpose ($p < .001$). Depressed mood yielded a larger effect size than *HDRS-17-sum* in 30 out of 32 instances ($p < .001$).

The pooled average effect size, as measured by *HDRS-17-sum*, was 0.27. There was approximately 30% greater separation between drug and placebo (ES: ~0.35) when a unidimensional subscale was used in place of *HDRS-17-sum* ($p < .001$). Effect sizes for individual items varied widely. *Depressed mood* (ES: 0.40) yielded approximately 50% larger drug-placebo differences than *HDRS-17-sum*, thus significantly outperforming all other items ($p < .001$), subscales ($p < .001$), as well as the *HDRS-17-sum* ($p < .001$).

Several items (*initial* and *middle insomnia*, *psychomotor agitation*, *somatic anxiety* and *insight*) yielded effect sizes ≤ 0.10, and for three individual items: *gastrointestinal symptoms*, *sexual symptoms* and *loss of weight*, effect sizes were negative, although significantly so only for *loss of weight* (p = .04). For these three items, it was more common that SSRI-treated patients had worsened as compared to baseline, than that placebo-treated patients had (OR 1.27, p = .009; OR 1.24, p = .005; OR = 1.21, p = .03).

3.1.3 Comment

The primary finding in this paper is that the effect of antidepressants is more consistent across trials than has usually been assumed, with only 9% of all included studies failing to indicate efficacy of the tested antidepressant when *depressed mood* was used as outcome parameter.

Of the three comparisons that were insignificant, two came from sertraline studies which used flexible dosing (50-200mg) and in which an active comparator (amitriptyline) had demonstrated assay sensitivity. The third came from a citalopram inpatient study in which the higher (40 mg), but not the lower dose (20 mg), demonstrated superiority over placebo, thus also demonstrating assay sensitivity. For citalopram this could be indicative of 20 mg being a less effective dose than 40 mg (see below, paper II). For sertraline it is unknown whether the two negative comparisons are a consequence of trial methodology or chance findings.

The average effect size of 0.27 for *HDRS-17-sum* is on par with previous reports.[130,149] That effect sizes were approximately 30% higher when using unidimensional subscales is similarly in line with previous findings.[89,90,231] While the unidimensional subscales thus showed significantly greater sensitivity for drug-placebo differences than the *HDRS-17-sum*, effect sizes for *depressed mood* (0.40) were significantly higher than for all subscales (p < .001), all other individual items (p < .001), and *HDRS-17-sum* (p < .001).

We did not exclude any studies based on duration, dosage, or design, apart from us enforcing a minimum sample size of 50 per arm. Hence, effect sizes are likely conservative.

Effect sizes for other symptoms varied widely, with some being relatively large whereas others were on average negative. Noteworthy is

that items reflecting insomnia and agitation did not contribute much to efficacy, hence refuting the theory that non-specific effects on these items, rather than a true antidepressant effect, are largely responsible for the separation seen between antidepressants and placebo.[72]

We found negative effect sizes at endpoint for three items that may be reflective of SSRI side effects: *gastrointestinal symptoms*, *sexual symptoms* and *loss of weight*. In a follow-up analysis of the frequency of worsening as compared to baseline, it was found that this was more common in SSRI-treated patients for all three items.

It should be noted that these results are likely conditional on the severity of the studied population. *Depressed mood* was highly prevalent, whereas, e.g., *suicidal ideation* and *psychomotor retardation* was far less common, and the latter symptoms were thus likely subject to floor effects which could serve to minimize the apparent beneficial effects of any effective treatment.

Our results thus suggest that the *HDRS-17-sum* is an amalgamation of several different effects, and that much information is lost when going from individual symptoms to the sum-score. That *depressed mood* was the most sensitive measure does not mean that it is the only informative measure, and we hence do not suggest that future trials in depression should use *depressed mood* as the sole outcome measure. What our results do illustrate is the large impact that the choice of rating instrument may have on apparent efficacy. And while decreased libido or weight loss may often be important symptoms of depression, it is problematic if they can also reflect side-effects of the active treatment. Not least for future drug development, the value of being able to separate a drug that is effective but which has side-effects, from one that is entirely ineffective, should be obvious.

In conclusion, in light of the many methodological problems inherent to conducting trials of antidepressants, the effect of SSRIs, as measured by *depressed mood*, was remarkably consistent across trials.

3.2 Paper II

3.2.1 Background

Systematic reviews and meta-analyses on the merits of dose-escalation in cases of non-response to SSRIs have concluded that this is not an effective practice. In addition, raising the dose may increase discontinuation symptoms and side-effects, as well as delay the recognition of early-state treatment-resistant depression.[80,232,233] It has thus been argued that this highly common practice should be abandoned, with the possible exception of for patients who are fast-metabolizers.[232]

Counter to this, recent meta-analyses of fixed-dose studies have concluded that there is evidence of dose-dependent efficacy. Papakostas and colleagues reported that higher starting doses of SSRIs were associated with significantly higher response rates, although the magnitude of the difference was not very large, and doses above double the usual starting doses (citalopram: >40mg, paroxetine: >40mg sertraline: >100mg) were not associated with increased efficacy.[234] In contrast, Jakubovski and co-workers reported dose-dependent beneficial effects of SSRIs up until doses exceeding those commonly recommended (citalopram: 60-75mg, paroxetine: 40-50mg, sertraline: 240–300mg).[233]

Relatedly, imaging studies have consistently found that usual SSRI starting doses produce roughly 80% occupancy at the serotonin reuptake transporter, and that higher doses do not significantly increase this.[235-238]

Of particular interest for this thesis is what influence the choice of outcome measure may have on the apparent dose-response relationship. Baker and colleagues have argued that, since side effects are dose-dependent,[80] any gain in efficacy may be offset by a larger proportion of dropouts in the higher dose-groups.[239,240] As shown in paper I, this could also affect the ratings of some adverse event-related HDRS items. Relatedly, if most antidepressant studies are barely powered to detect the HDRS sum-score difference one might expect between a placebo and an effective dose of an antidepressant, they are likely underpowered to detect differences between two antidepressant doses which both outperform placebo.[84,241]

The clinical importance of this issue is evident: if high doses are no more effective than lower ones, the commonly applied strategy of dose-escalation will be harmful to patients by increasing the burden of side-effects, with no corresponding gain in efficacy.[232] On the other hand, if they do display dose-dependent beneficial effects, the supposition that they do not may instead influence clinicians to use suboptimal doses.[242] And yet another reason to clarify this issue is that the alleged lack of dose-response relationship has been put forward as an argument by debaters questioning the usefulness of these drugs; a common claim thus has been that a truly effective drug *should* display a dose-response relationship, and that the absence of such a relationship for the SSRIs suggests them to be ineffective.[157,163,243]

In paper II[244] we hence investigated whether or not the beneficial effects of SSRIs in fixed dose-studies (11 trials, n = 2 859) are dependent on dose. We hypothesized that a possible beneficial dose-response relationship on core depression symptoms may be obfuscated by a larger detrimental effect on items likely to be sensitive to SSRI side-effects in those receiving higher doses, and that assessing *depressed mood* therefore might be more informative than assessing *HDRS-17-sum*.

3.2.2 Results

An initial analysis stratified by drug (citalopram, paroxetine, sertraline) offered indications of dose-dependent beneficial effects. This appeared to be primarily due to doses in the subtherapeutic to low therapeutic range (citalopram: 10 and 20mg, paroxetine: 10mg, sertraline: 50mg) being less effective than doses in the low to mid therapeutic range (citalopram: 40 and 60mg, paroxetine: 20-40mg, sertraline: 100 and 200mg).

We thus pooled all patients and subdivided the SSRI population into a *low-dose* group and an *optimal-dose* group, based on the abovementioned cut-offs.

Low-dose significantly outperformed placebo, but was significantly less effective than *optimal-dose* with regards to change in *depressed mood* (*low-dose* vs placebo ES = 0.30; *optimal-dose* vs placebo ES = 0.51) and *HDRS-17-sum* (*low-dose* vs placebo ES 0.19; *optimal-dose* vs placebo ES = 0.36). Corresponding differences were seen on the categorical outcomes response and remission.

Acknowledging that the procedure of visually inspecting the results for individual drugs and then selecting cut-offs may be questioned we conducted two sensitivity analyses: *i)* an 'extreme dose' analysis which only included the lowest and highest dose administered for each drug and *ii)* an analysis with dose treated as a linear predictor. Placebo-treated patients were excluded from the latter analysis. Results from both sensitivity analyses mirrored those of the primary one; the 'extreme dose' analysis however yielded significant differences between effective doses only with respect to depressed mood (ES: 0.26, $p < .01$) but not with respect to *HDRS-17-sum* (ES: 0.15, p = ns).

We also conducted a linear sensitivity analysis of doses within the *optimal-dose* group. This revealed no indication of a dose-response relationship for either *depressed mood* ($p = .5$) or *HDRS-17-sum* ($p = .3$) within this group.

Incidentally, the use of longitudinal modelling revealed that while differences as measured by *HDRS-17-sum* were generally not observed before three weeks of treatment, there were small but significant differences in favour of SSRI-treatment already after one week with respect to *depressed mood* (ES range: 0.11-0.23).

3.2.3 Comment

In line with some,[233,234] but not all,[243] previous meta-analyses we did observe an influence of dose on efficacy.[157,163] Consequently, average effect sizes for *optimal doses* were somewhat larger than what has been commonly reported in SSRI meta-analyses for *HDRS-17-sum* (ES 0.34 to 0.36), and much larger for *depressed mood* (ES 0.51 to 0.57). We suggest that the beneficial effects of SSRIs in MDD have been underestimated by meta-analyses failing to account for *i)* the problems associated with the *HDRS-17-sum* as effect parameter and *ii)* dose.

Similar to Papakostas and co-workers,[234] but in contrast to Jakubovski and colleagues,[233] differences between effective doses were only found between doses in the subtherapeutic to low range on the one hand, and doses in the medium-range on the other hand. We found no added benefit of doses that may be described as high (citalopram 60mg, paroxetine ≥ 30mg, sertraline 200mg).

We suggest that the comparatively high dose-ranges for maximum beneficial effects arrived at by Jakubovski and associates are due to the way they included and characterized flexible-dose studies.[245] By treating flexible dose-studies as if all included participants received the maximum dose these studies are bound to cluster at the top of all included studies with respect to dosage. If flexible-dose studies also yield larger average effect sizes than fixed-dose studies (which they have been reported to do),[246] this would confound effects in such a way as to make very high doses seem the most effective.

Notably, our results are compatible with those from the dose-escalation literature which have failed to demonstrate any benefit from this practice. Thus, the doses from which these studies have begun escalating (paroxetine 20mg and sertraline 100mg)[238,247,248] already produce, on average, maximum benefit in our analyses. The one, small (n=91) dose-escalation study that may be informative with regards to our results, is one by Schweizer and co-workers that investigated dose-escalation from 50 to 150mg of sertraline.[249] While not achieving statistical significance, the direction and magnitude of the results were in agreement with superior efficacy of 150mg compared to 50mg (*HDRS-17-sum* difference of 2 points).

While differences between active drug and placebo with respect to *HDRS-17-sum* were almost invariably non-significant during the first two weeks of treatment, there were consistent improvements in *depressed mood* already from the first week in most analyses. Similar findings have been reported previously,[250] and in conjunction with reports of negative efficacy for certain HDRS items early in treatment,[251] they suggest that perhaps antidepressant efficacy with regards to mood is not so much delayed in onset as it is gradual.

3.3 Paper III

3.3.1 Background

The fact that inert placebos do not produce any side effects, whereas almost all pharmacologically active substances do, poses a problem when conducting double-blind trials.[252,253] It could bias the observers, who, noticing recognizable side-effects, might be influenced to rate outcomes differently.[252] And it could influence patients, who, having been briefed on the same side-effects, may conclude that they have been randomized to active treatment; a realization which, it has been hypothesized, could introduce an amplified placebo effect.[54,162,163,254] The latter theory has been widely disseminated both in lay media and the scientific literature as an alternative explanation for why antidepressants are superior to placebo in clinical trials.[156-158,163,170]

While studies have suggested that clinicians and patients are better than chance at guessing treatment assignments,[255] it is not obvious what influence this may have on efficacy ratings. Moreover, for any effective treatment one could get a better than chance prediction by simply guessing active treatment whenever there is improvement.[256,257]

In the 1960s and 1970s, attempts were made to control for this potential confounder through the use of active placebos. Specifically, the anticholinergic agent atropine, which mimics many side effects of tricyclic agents, was used instead of inert placebo.[254,258] These active placebo trials were reviewed by Moncrieff and co-workers[258] who reported an overall effect size of 0.39 (0.24 to 0.54), and a conservatively estimated effect size of 0.17 (0.00 to 0.34) when one positive trial which was deemed an outlier was excluded, concluded that: '*this suggests that unblinding effects may inflate the efficacy of antidepressants in trials using inert placebos*.'

Quitkin and colleagues reviewed the same collection of trials and noted that they suffered from important drawbacks common to many trials of the era: inadequate doses, low power, short treatment duration, and diagnostic heterogeneity.[259] Further, they suggested that if it were true that active placebos induce a stronger placebo effect via unblinding, then one would expect active placebo groups to have higher response rates than inert placebo groups. The average placebo response rate in these active placebo trials being 22%, which is on the lower end of what was commonly seen in antidepressant trials with inert placebos at the time,[54]

hence does not support the hypothesis that the use of active placebo induces an amplified placebo response.

In support of a significant role for unblinding effects has also been presented early trial-level meta-analyses showing the frequency of side-effects to correlate with antidepressant efficacy.[260] These results, however, have failed to replicate in another and much larger trial-based analysis,[261] and both of these efforts are of course also susceptible to the ecological fallacy.[228,262]

Another piece of supporting evidence for the 'side effects breaking the blind'-hypothesis is that response to treatment is larger in trials that only include active comparators than it is in placebo-controlled trials.[263] Also those observations are, however, difficult to interpret due to trials using active comparators also displaying lower rates of dropout than do placebo-controlled trials.[263] On the other hand, arguing against this hypothesis is the fact that differences in efficacy have been seen between active treatments in trials that lack a placebo arm, i.e., under conditions where this putative bias should not be able to be expressed.[264-267]

The importance of having this controversy resolved is of obvious clinical importance. If the efficacy of antidepressants is indeed artefactual and driven merely by unblinding, the extensive use of SSRIs for depression, and possibly other indications as well, should be reconsidered.[228] If this theory is incorrect, however, it is equally important to have that clarified so that doctors are not discouraged from prescribing effective medication. Assessing this is thus the subject of paper III,[268] in which we contrasted patients receiving treatment (citalopram or paroxetine) against those given placebo conditional on whether or not they reported side effects.

3.3.2 Results

Patients treated with paroxetine or citalopram fared significantly better than patients treated with placebo regardless of whether they reported an early (week 1 or 2) side effect (paroxetine ES: 0.48, $p < .001$; citalopram ES: 0.31, $p = .002$) or not (paroxetine ES: 0.33, $p < .001$; citalopram ES: 0.49, $p < .001$). These results were replicated in sensitivity analyses *i)* using alternative time-periods for determining side-effect status (week 1 or 2 + lead-in, week 1, week 1 to 6), *ii)* including only adverse events deemed probably or possibly related to

treatment, *iii)* stratifying also the placebo-group according to whether or not they reported early adverse events, and *iv)* using the intention-to-treat rather than the observed cases population. For paroxetine, the results were also similar in a sensitivity analysis using *HDRS-17-sum* as outcome measure rather than *depressed mood,* whereas for citalopram, patients with early adverse events did not show significant separation from placebo-treated patients in this analysis (ES: 0.17, p = .10).

Paroxetine-treated patients with side-effects had significantly lower outcome ratings than those without (ES: 0.15, p = .008). This was not the case for citalopram-treated patients where instead those without side-effects had lower outcome ratings than those with side-effects (ES: -0.17, p = .14). These findings were mostly consistent across sensitivity analyses, although when using *HDRS-17-sum* as outcome parameter there was no significant difference between paroxetine patients with and without side-effects (ES 0.07, p = .20).

Adverse event severity did not moderate treatment response for either paroxetine (p = .93) or citalopram (p = .48).

3.3.3 Comment

Patients treated with the SSRIs paroxetine and citalopram fared significantly better than patients treated with placebo regardless of whether or not they reported side effects. This observation is not compatible with the hypothesis that side-effects are largely or entirely responsible for the efficacy of antidepressants.[157,158,162,163]

On the other hand, paroxetine-treated patients with side-effects demonstrated a small but significant improvement over paroxetine-treated patients without side-effects when efficacy was measured by *depressed mood*. While this is in line with a potential role for unblinding influencing efficacy, there are obviously many other possible reasons for why beneficial and adverse effects may correlate (e.g., through differences in dose, compliance, drug metabolism, or pharmacological responsiveness). While not significant, the opposite tendency was observed for citalopram.

One limitation that should be commented on is the possibility that subthreshold side-effects, which were noticeable to the patient, but not of sufficient severity to merit an adverse event report, made it so that

also patients who did not report side-effects believed they had been given active treatment. Considering the high prevalence of side effects in this material, and that previous studies have found far from a 100% correct classification rate,[255] we however deem this an unlikely confounder. Relatedly, adverse event severity did not moderate efficacy for either drug.

3.4 Paper IV

3.4.1 Background

Whether antidepressants may provoke suicidality in susceptible individuals has long been a controversial issue (see 1.3.1). In paper IV[269] we look at rating scale-assessed suicidality as measured by the HDRS item *suicidal ideation*.

We analysed both group mean ratings of *suicidal ideation* and the incidence of three categorical derivatives of the *suicidal ideation* item. These were: *i)* worsening, defined as a higher rating of *suicidal ideation* than at baseline, *ii)* emergent suicidality (loose), defined as a *suicidal ideation* rating of 2 to 4 in a patient who scored 0 or 1 at baseline,[270] and *iii)* emergent suicidality (strict), defined as a *suicidal ideation* rating of 3 or 4 in a patient who scored 0 or 1 at baseline.[181] The purpose of the former analyses was to assess the net impact of treatment, whereas that of the latter was to identify possible negative effects in a subgroup of patients.

Prompted by earlier reports suggesting the influence of SSRIs on suicidality to be age-dependent, all analyses were stratified by age, with young adults ($18 \leq age < 25$) and adults ($age \geq 25$) analysed separately.

3.4.2 Results

For adults, mean ratings of *suicidal ideation* were significantly improved after one week of treatment (ES: 0.09, $p < .001$) and remained so throughout the study (endpoint ES: 0.25, $p < .001$). Sensitivity analyses of the ITT population (ES: 0.20, $p < .001$), and of patients that dropped out (ES: 0.23, $p < .001$), yielded comparable results.

For young adults, mean ratings of *suicidal ideation* were significantly lower in SSRI-treated subjects after one week of treatment (ES: 0.23, p = .03), but there were no significant differences at subsequent visits, nor in the ITT population or among patients who dropped out.

Follow-up analyses revealed young adults to score slightly but significantly higher on *suicidal ideation* at baseline (ES: -0.15, p = .001). An interaction analysis between the age groups found that while there were no differences in the response to SSRI between them, placebo-treated young adults responded significantly better than placebo-treated adults (interaction p = .009).

Kaplan-Meier analyses found SSRI treatment to significantly decrease the risk of worsening (p < .001) and emergent suicidality (loose) (p < .001) in adults, but showed no significant effect on emergent suicidality (strict). These results were mirrored by logistic regression analyses looking separately at early (week 1 and 2) and late (week 3 to 6) occurrences of deterioration.

For young adults, Kaplan-Meier analyses found no significant differences on any categorical measure of deterioration. These results were mostly replicated in the logistic regression analyses; however, worsening (p = .04) and emergent suicidality (loose) (p = .01) were significantly more common in SSRI-treated patients during the late, but not the early, study period.

3.4.3 Comment

For adults, effects on the HDRS *suicidal ideation* item were consistently in favour of SSRI-treated patients. This was evident both in analyses of group mean effects, and in categorical analyses of three measures of deterioration.

In contrast, the results for young adults were inconclusive, with most analyses not detecting any significant effect of SSRI treatment. The exception being the logistic regression analyses of worsening and emergent suicidality (loose), which were both significantly more common in SSRI-treated young adults during the late study period (week 3 to 6) but not during the early study period (week 1 and 2), or in patients who dropped out. Similar effects were not seen in the Kaplan-Meier analyses which included the entire study period (week 1 to 6).

While notable, in consideration of the large number of tests conducted, the relatively weak significances, and the small sample size of the young adult group (n=414 for worsening, n=247 for emergent suicidality, loose), these findings should be interpreted with caution.

While our results suggest the effects of SSRIs on suicidality to be decidedly beneficial in adults, it cannot be excluded that certain adult patients experienced a drug-induced deterioration, but that this was not possible to detect due to a larger fraction of patients who improved with SSRI treatment.

In summary, our results mirror those from a previous publication concerning fluoxetine and venlafaxine,[193] which also found HDRS-rated *suicidal ideation* to improve in adults treated with these substances, whereas the effect in young adults was inconclusive. Together, these findings suggest that it is highly unlikely that the SSRIs are associated with a large suicide-provoking effect. This conclusion is in line with several reports showing an inverse relation between SSRI prescriptions and suicide rates.[123,217-219]

3.5 Paper V

3.5.1 Background

A common,[55,149,190,271-277] but not universal,[278,279] meta-analytical finding has been a tendency for drug-placebo differences to increase with depression severity. This, in conjunction with the recent emphasis on clinical significance,[145,149,190,276] has resulted in regulatory agencies and expert bodies dissuading from the use of pharmacological antidepressants for non-severe cases.

Two highly influential publications in this debate are the 2008 meta-analysis in PLOS Medicine by Kirsch and colleagues,[149] and a comparatively small (n=718) mega-analysis published in JAMA two years later by Fournier and co-workers.[276] Both of these reported results in support of increased drug-placebo separation with increasing baseline severity.

Since the publication of these reports, several large-scale investigations addressing the possible severity-by-treatment interaction using patient-

level data have been conducted.[280] Most[281-283] but not all[284] of these have failed to find statistical support for such an interaction to be at hand. Of note is the study by Rabinowitz and co-workers who, while not finding support for such a severity-by-treatment interaction when analysing data on the patient-level, did find a significant interaction when analysing the same data using study-level aggregates. Conversely, Petkova and co-workers found such an association when analysing the data on patient-level, but not when meta-regressing aggregate study-level data.[282]

Of note is that most meta-analyses, and all five mega-analyses, have used the HDRS for assessing outcome and severity. As mentioned in the introduction, and in the chapter presenting the results of paper I, there are however reasons to doubt the fitness of this instrument for both these purposes.

If the conclusion of the oft-cited meta-analysis by Kirsch and co-workers[149] on this topic, i.e. that antidepressants produce a meaningful improvement (if at all) only in the '*the most extremely depressed patients*', is correct, then the current use of these drugs must obviously be reconsidered. We hence deemed it important to address this issue using patient level data for citalopram, paroxetine and sertraline, our addition in comparison to previous patient-level analyses being to assess the impact of baseline severity not only on *HDRS-17-sum*, but also on the following HDRS-derived outcome measures: *depressed mood*, *HDRS-6-sum*,[70,71] and *non-HDRS-6-sum* (i.e. the sum of the 11 HDRS items not included in *HDRS-6-sum*)

3.5.2 Results

We found no severity-by-treatment interaction when both outcome and initial severity were measured by *HDRS-17-sum* in the full population. Visual inspection of the plots, however, hinted that there may nevertheless be a linear severity-treatment interaction that was obscured by high variance in the extreme ranges of the severity distribution. To assess this, we reran the analysis after having excluded severity values with less than 50 observations per treatment arm; resulting in an HDRS range of 18 to 29 (92% of all patients). This manoeuvre revealed a significant interaction in both the MMRM ($\beta = 0.14$ (0.067), $p = .036$) and the LOCF-analyses ($\beta = 0.16$ (0.061), $p = .007$); likewise, similar results were obtained from yet a sensitivity analysis in which we instead

excluded severity values with less than 20 observations per treatment arm (HDRS range 17 to 32; 97% of all patients) (MMRM p = .044; LOCF p = .016)

Importantly, analyses using *depressed mood* or *HDRS-6-sum* as outcome parameters did not find significant severity-by-treatment interactions, regardless of whether uncommon severity values were included or not. Conversely, when *non-HDRS-6-sum* was used as measure of outcome, the severity-treatment interaction was significant across all three populations. Using *HDRS-6-sum* to assess baseline severity and *HDRS-17-sum* to measure outcome also resulted in non-significant interactions.

Partitioning patients into those with severe (*HDRS-17-sum* $\geq$ *27*) and non-severe (*HDRS-17-sum* $\leq$ 18) depression yielded comparable results: *non-HDRS-6-sum* again being the only outcome measure which showed a significant interaction (p = .005).

Exploratory analyses contrasting the non-severe category against the severe category found that while symptoms belonging to the *HDRS-6* cluster were highly prevalent in both severity groups, symptoms belonging to the *non-HDRS-6* cluster were markedly less common in non-severe patients. Thus, while mean *HDRS-6-sum* ratings differed by 37% between groups, *non-HDRS-6-sum* pathology was 113% higher in the severe group. Correspondingly, effect sizes for *HDRS-6* symptoms were roughly in parity between the two groups, with many *non-HDRS-6* items showing negative or null effect sizes in non-severe patients, but positive ones in severe patients.

3.5.3 Comment

A major conclusion of this study is that also subjects with relatively low symptom rating at baseline experience a marked reduction in core symptoms of depression when treated with an SSRI. The widespread assumption that they are useless for the treatment of non-severe depression hence seems unfounded.

That we did not detect an overall interaction between severity and treatment as measured by *HDRS-17-sum* in the full population is in line with most previous mega-analyses.[193,282,283] Notably though, all three of these displayed non-significant tendencies in favour of greater drug

efficacy with increasing baseline severity. We hence suggest that such an interaction is real, but sometimes, as in our analysis, masked by rare, extreme values.

There are many factors that may explain the finding that the reduction in non-core symptoms, but not in core symptoms, seems dependent on baseline severity. First, the items belonging to *HDRS-6* were much more common than other symptoms in non-severe depression. This is not unexpected as commonness was one of the criteria used when constructing this subscale, [71] and also since many studies in our material have used a minimum cut-off on *depressed mood* as an inclusion criterion thus guaranteeing its presence. That most patients do display these symptoms likely alleviates floor effects on items for which the SSRIs display strong beneficial effects.

Second, many of the items belonging to the *non-HDRS-6* cluster have been shown to display minute or even negative effect sizes, and may reflect common SSRI side-effects.[229,251] For patients experiencing such symptoms with high severity at baseline, there may be a ceiling effect in the sense that the rating instrument does not permit further deterioration; moreover, patients with, e.g., severe sexual dysfunction or somatic anxiety already at baseline may experience a relief with regard to these symptoms even in the presence of SSRI induced side-effects, provided that their depression improves significantly. In contrast, for subjects with mild depression, in which these symptoms are often rare or mild at baseline, the impact of SSRI side-effects may be considerably larger. Relatedly, many *non-HDRS-6* items (e.g., *insomnia, somatic anxiety*, and *gastrointestinal symptoms*) are common also in healthy people, and may thus be less responsive to an overall change in the severity of depression.

Third, a complementary or additional explanation could be that SSRIs are only effective in a fraction of patients, and that the size of that fraction is independent of baseline severity. This has been indicated in one large meta-analysis.[278] If so, and if the average improvement in this fraction is a constant percentage of baseline symptomology, then one would expect to see larger numerical separation with increasing baseline severity. That we did not observe any indications of a severity-treatment interaction when using *HDRS-6-sum* as measure of both baseline severity and outcome (data not shown), however, argues against this being the entire explanation.

Notably, the two mega-analyses that did detect a significant interaction between baseline severity and treatment both included trials of imipramine together with a selective serotonin reuptake inhibitor (paroxetine and fluoxetine, respectively), whereas the other three did not include any TCA.[276,284] Imipramine being associated with a larger burden of side-effects[285] and/or with better efficacy on sleep-related items,[286] could perhaps explain why these two mega-analyses found evidence for a severity-treatment interaction, while the other three did not.

Additionally, as all HDRS scale items are measured on the ordinal level, it is likely that precise relations will be non-linear and conditional not only on which symptom is measured, but also on its exact rating, as some HDRS symptom ratings may be more sensitive to change than others.

Four limitations deserve discussion: *i)* patients in the extreme group analyses (non-severe and severe, respectively) differ not only in severity on each specific item but also with respect to overall pathology. A possible confounder is that patients with lower overall severity may be more prone to experience side-effects as detrimental, potentially leading to dropout. *ii)* The precise division between *HDRS-6* and *non-HDRS-6* is arbitrary. While the existence of a core depression factor has been replicated multiple times,[65] there is no consensus on precisely which items should be included. For example, Santen and colleagues proposed a seven-item scale which included the items in the *HDRS-6* as well as *suicidal ideation*,[91] and the Toronto scale proposed by McIntyre and co-workers include *somatic anxiety* and *suicide* but not *psychomotor retardation.*[287] *iii) Somatic anxiety* and *insomnia* are more common in unipolar than bipolar depression.[270] Thus patients with low ratings on these symptoms may be more likely to suffer from undiagnosed bipolar disorder, which has been reported to respond less favourably to antidepressant treatment.[288] *iv*) Our material contained very few patients with *HDRS-17-sum* lower than 17, i.e., very few patients had truly mild depression and our results can therefore not inform the debate on to what extent such subjects may respond to SSRIs.[55,84]

In summary, when looking at core symptoms of depression there is no influence of baseline severity and patients display meaningful benefits from antidepressant treatment regardless of severity. Thus, while *HDRS-17-sum* differences do increase with baseline severity, this appears to be driven by specific symptoms not included in the *HDRS-6*.

These symptoms not responding in patients categorized as non-severe may partly be explained by them being uncommon in patients with low overall severity; these symptoms thus account for the lion's share of the *HDRS-17-sum* increase across the severity range that has commonly been included in antidepressant trials.

3.6 Paper VI

3.6.1 Background

All of the previous papers detail results from post-hoc analyses stemming from one data set (or subsets thereof) in which drugs belonging to the SSRI class have been administered. One major aim of paper VI was thus to assess to what extent the findings obtained for the SSRIs could be replicated in a population of studies concerning an antidepressant with a somewhat different profile: the SNRI duloxetine.

It has previously been suggested that duloxetine, because of its affinity for the noradrenaline transporter, displays a somewhat different effect profile as compared to the SSRIs, including a more pronounced effect on the items *work and interests* and *psychomotor retardation*.[289] A second aim was to address this possibility by comparing effect profiles for all HDRS items (after adjusting for overall response) in the duloxetine population and in the SSRI population, respectively.

3.6.2 Results

Whereas the individual items *depressed mood* (ES: 0.20), *guilt* (ES: 0.10), *suicidal ideation* (ES: 0.15), *agitation* (ES: 0.08) and *psychic anxiety* (ES: 0.23) separated significantly and positively when comparing duloxetine with placebo already at the week 1 evaluation, *HDRS-17-sum* (ES: 0.05, $p = .16$) did not. The latter observation was explained by the ratings of *middle insomnia* (ES: -0.08), *gastrointestinal symptoms* (ES: -0.30), *sexual dysfunction* (ES: -0.10), and *loss of weight* (ES: -0.22) showing significant initial deterioration in patients treated with duloxetine.

At endpoint, all individual symptoms except *middle insomnia*, *loss of weight* and *insight* significantly favoured duloxetine over placebo. *Depressed mood* (ES: 0.44) had the highest effect size, closely followed by *psychic anxiety* (ES: 0.43), whereas *HDRS-17-sum* was slightly lower (ES 0.37). For all other items, effect sizes ranged from -0.05 (*loss of weight*) to 0.32 (*work and activities*).

Separate age-stratified analyses of the *suicidal ideation* item revealed significant superiority of duloxetine over placebo from week 1 until week 8 for patients aged 25 and over (n=3318), and non-significant differences numerically in favour of duloxetine in patients aged below 25 (n=257).

In paper IV we reported that some categorical measures of aggravated suicidal ideation were more common in SSRI-treated young adults than they were in placebo-treated young adults during the late study period (week 3 to 6), but not during the early study period (week 1 and 2). While the corresponding categorical analyses were deemed to be beyond the scope of the report regarding duloxetine, and are hence not included in manuscript VI, results from analyses corresponding to those in paper IV, are presented below.

As our previous findings concerned SSRI-treated patients, we first contrasted SSRI-treated young adults to placebo-treated young adults with respect to the incidence of worsening and emergent suicidality (loose and strict definition, respectively) during the late study period. Worsening thus occurred in 10 out of 47 young adults treated with placebo, compared to 7 out of 46 treated with an SSRI. Emergent suicidality (loose definition), occurred in 4 out of 46 young adults treated with placebo, and 4 out 43 treated with SSRIs. One placebo-treated patient and no SSRI-treated patient reported emergent suicidality according to the stricter definition. Due to the low number of patients and events, we refrained from conducting any statistical tests.

To increase power, we also conducted an analysis in which SSRI- and duloxetine-treated young adults were pooled. This yielded non-significant differences numerically favouring active treatment. Worsening thus occurred in 23 out of 161 placebo-treated young adults as compared to 17 out of 163 actively treated patients (OR 0.70, 0.36 to 1.37, p = .29), emergent suicidality (loose definition) occurred in 13 out of 145 placebo-treated patients and 10 out of 138 actively treated patients (OR 0.79, 0.34 to 1.88, p = .60), and there was one occurrence

of emergent suicidality according to the strict definition in each group (OR 1.05, 0.06 to 17.2, p = .97).

There were no significant interactions between baseline severity (as measured by *HDRS-17-sum*) and treatment when outcome was assessed by *depressed mood*, *HDRS-6-sum* or *HDRS-17-sum*. When outcome was assessed by *non-HDRS-6-sum* there was, on the other hand, a significant interaction in the direction of larger drug-placebo differences with increasing *HDRS-17-sum* severity (β = 0.08 (0.03), p = .02). As in paper V, this was driven by low effect sizes for non-core symptoms in patients categorized as non-severe (*HDRS-17-sum* ≤ 18). Also in line with paper V, the prevalence and severity of non-core symptoms differed markedly between cases classified as severe (*HDRS-17-sum* ≥ 27) and non-severe. Mean *HDRS-6-sum* ratings thus were merely 63% higher in severe cases than in non-severe cases, whereas *non-HDRS-6-sum* ratings were 139% higher.

Duloxetine-treated patients fared significantly better than placebo-treated patients with regard to *depressed* mood, irrespective of whether they reported early (week 1 to 2) adverse events (ES: 0.46) or not (ES: 0.34). The within-treatment comparison however significantly favored duloxetine-treated patients with early adverse events (ES: 0.12, p = .02). Similar results were obtained when including adverse events occurring during the entire study period (week 1 to 8). When using *HDRS-17-sum* as measure of efficacy, results were similar in that both duloxetine groups outperformed placebo; the within-treatment contrast, however, was non-significant. Adverse event severity did not significantly moderate outcome as measured by *depressed mood*.

The analysis of efficacy-adjusted symptom response profiles yielded strikingly similar results for duloxetine and the SSRIs. There were hence no significant differences between treatments for any HDRS item.

3.6.3 Comment

In line with paper II, *depressed mood* significantly outperformed placebo already after one week of treatment whereas *HDRS-17-sum* did not.[244] Also, consistent with paper I, *depressed mood* displayed a larger effect size at endpoint than did *HDRS-17-sum*.[229] The magnitude of this difference (24%) however was smaller than that for the SSRI population (48%) at the corresponding time-point (week 6).

Similarly, the finding of early and enduring positive effects on rating scale-assessed *suicidal ideation* for patients aged 25 or above is in line with results from paper IV, as are the non-significant differences seen in patients below the age of 25; all differences were however numerically in favour of duloxetine in the young adults.[269] Similarly, while not included in manuscript VI, we did not observe any indications that SSRI or duloxetine treatment caused increased suicidality in young adults.

The finding of a significant severity-by-treatment interaction when outcome was assessed by *non-HDRS-6-sum*, but not when using other outcome measures (*depressed mood*, *HDRS-6-sum*, *HDRS-17-sum*), mirrors results from the SSRI population (paper V), as does the observation that differences in the rating of non-core rather than core symptoms of depression constitute the major difference between patients categorized as severe and non-severe, respectively. The assumption that the tendency for baseline severity to be positively associated with response to a large extent may be due to the HDRS comprising items rated low at baseline in cases categorized as non-severe hence gained further support from the duloxetine dataset.

That duloxetine-treated patients outperform placebo regardless of the presence of adverse events, and that efficacy is not moderated by adverse event severity, is in line with the findings in paper III.[268] The results in paper III differed between SSRIs (citalopram and paroxetine) as to the strength and direction of the possible association between adverse events and efficacy as measured by *depressed mood*. Duloxetine-treated patients with side-effects, like paroxetine-treated patients with side-effects, fared slightly but significantly better than patients without side-effects. As discussed for paper III, while compatible with the side-effects breaking the blind hypothesis, there are several other explanations for why such an association may be seen.

On the basis of seven head-to-head comparisons of duloxetine versus different SSRIs, it has previously been suggested that duloxetine exerts a more pronounced effect on, e.g., psychomotor retardation and inability to work, than do the SSRIs;[289] the interpretation of these trials is however complicated by the use of relatively low and possibly suboptimal SSRI doses in several of them. Our analysis of symptom response profiles lent no support for any major differences between duloxetine and the SSRIs, but revealed strikingly similar profiles for the two treatments. While adequately powered head-to-head analyses using equipotent doses are always required to definitively settle the issue of

possible differences in symptom-level efficacy between two treatments, our analysis suggest duloxetine to be highly similar to the SSRIs with regards to symptom-level response pattern.

In summary, the findings from paper I to V translated remarkably well to an independent population of duloxetine studies. In line with this, our item-level analyses of efficacy suggest the response profile of duloxetine to be highly similar to that of the SSRIs.

4 Conclusion

The findings in this thesis contradict many of the arguments put forth as evidence of SSRIs being ineffective, or exerting merely non-specific effects, for the treatment of MDD. The largest effect of SSRI treatment was an improvement in mood, not a non-specific improvement. This was a consistent finding, with 91% of all drug-placebo comparisons displaying significant improvement on this symptom with SSRI treatment. Likewise, the effect of SSRIs on suicidal ideation was clearly beneficial in patients above the age of 25. Both of these effects were evident already after one week of treatment.

The beneficial effects of SSRI treatment, rather than (as often claimed) being unrelated to dose, were dose-dependent, although maximum efficacy was achieved already at comparatively low doses. Similarly, side-effects, rather than being solely or largely responsible for the drug-placebo differences observed in antidepressant trials (as often claimed), turned out not to be necessary for SSRIs to outperform placebo. Further, rather than being clinically meaningful only for the most extremely depressed patients (as also often claimed), the effects of SSRIs on core depression symptoms were as large for the least severe patients as they were for the most severely depressed. It should however be considered that subjects with truly mild depression are seldom included in trials of antidepressants.

Most of these findings come as a result of abandoning the outcome parameter that was originally used in the included trials, i.e., the HDRS sum-score. We conclude that this parameter is an insensitive measure of SSRI efficacy, in part because it captures SSRI side effects. This was most apparent early in treatment where the HDRS sum-score consistently found no differences, whereas item-level analyses showed clear effects on many symptoms. While side effects may impact quality of life and should be meticulously recorded, they should preferably not affect the rating of depressive severity as this will effectively amount to double-counting the adverse events. It will also render it more difficult to assess the magnitude of the beneficial effects of SSRI treatment.

In summary, the use of an insensitive effect parameter has resulted in an underestimation of the value of SSRIs for the treatment of MDD, and

may also have impeded drug development in depression by making drug candidates appear less effective than they really are. While we would thus recommend that one in future trials in depression refrains from using the *HDRS-17-sum* as primary effect parameter, we note with appreciation that modern trials indeed appear to be applying other outcome measures, e.g., the MADRS, to an increasing extent.[144,290,291]

5 Methods and materials

5.1 Ethics

Secondary use of de-identified patient-level data does not require ethical approval in Sweden. We have however obtained an advisory opinion from the Regional Ethical Review Board of Gothenburg, Sweden, in which they state no objection to the conduct of these studies.

5.2 Statistical analyses

Continuous outcome measures have been analysed using analysis of covariance (ANCOVA), either for the ITT- or the OC-population, and/or using mixed models for repeated measurement (MMRM).[142] Unless expressly stated otherwise, all models have included a fixed factor for centre or trial (depending on the level of analysis), and baseline severity as a covariate. MMRM models have also included a fixed factor for time (visit or week) and the interaction between time and treatment. MMRM models have been specified to use unstructured (co)variance matrices to model within-patient errors and the Kenward-Roger approximation to estimate denominator degrees of freedom.[292] Effect sizes for continuous measures have been calculated by dividing least-squares means differences by the root of the corresponding variance term. Effect estimates for categorical measures are generally reported as odds-ratios. ITT-populations have been constructed using the LOCF procedure.

Categorical outcomes have been assessed using logistic regression, generalized linear mixed models (GLMM), and Kaplan-Meier analyses. Logistic regression models have used corresponding specifications to the ANCOVA models, whereas GLMM-models have been specified in an analogous fashion to the MMRM-models. GLMM models used for dichotomous outcomes have used a binary distribution with a logit link.

In paper I, we investigated SSRI efficacy, regardless of dose, using LOCF ANCOVA. Hence fixed factors for treatment (placebo, SSRI) were included in the analyses. Treatment-by-centre interactions were assessed for all individual studies, and significant treatment-by-centre interactions were handled through sequential exclusion of the centres

displaying the most divergent outcomes. The choice of using LOCF ANCOVA rather than more advanced modelling (e.g., MMRM) was motivated by this having been the most commonly used analytical procedure when these trials were originally reported. We hence attempted to replicate the original analyses as closely as possible. Ordinal logistic regression was used to ensure that the use of ANCOVA to analyse the ordinal measure *depressed mood* did not impact the results. Repeated measures analysis of variance and paired t-tests were used to test for differences between effect parameters. McNemar's test was used to test whether the ratios of comparisons reaching statistical significance differed between *depressed mood* and *HDRS-17-sum*. Pearson's one-sample $\chi 2$-test was used to assess whether the ratio of effect size estimates which improved when switching to *depressed mood* was significantly different from what can be expected by chance.

In paper II, SSRI dose was included both as a categorical (placebo = 0, actual doses in mg) and as a continuous predictor in MMRM-models. When treated as a continuous predictor, doses were normalized so that the lowest included dose (for each drug) was anchored at zero, the highest at one, with doses in between linearly interpolated. Placebo-treated patients were excluded from all analyses treating dose as a continuous predictor. GLMM models with dose group as a fixed factor were used to assess the categorical outcomes remission (*HDRS-17-sum* ≤ 7, *depressed mood* = 0) and response (*HDRS-17-sum* reduction of at least 50%, *depressed mood* decrease of at least 1 point).

In paper III, we wished to investigate the relation between side-effects and treatment outcomes. As side effects may lead to treatment discontinuation, we considered the OC-population as primary. These analyses included a three-level fixed factor to classify patients (placebo, SSRI without adverse events, SSRI with adverse events). When assessing whether side-effect severity moderated outcomes, patients were stratified according to the highest severity side effect they had experienced (mild, moderate, severe). Patients without adverse events, and patients treated with placebo, were excluded from these analyses.

In paper IV, the impact of SSRI treatment on HDRS assessed suicidality was assessed using LOCF ANCOVA and MMRM. We also conducted analyses encompassing only those patients that dropped out prior to study end. All models included SSRI treatment as a fixed-factor (placebo, SSRI). Categorical measures of deterioration were analysed using both logistic regression and Kaplan-Meier survival analysis.

Patients who by definition could not experience deterioration on a particular measure were excluded from the corresponding analyses.

In paper V we assessed the relation between baseline severity and efficacy. LOCF ANCOVA and MMRM models were both used. Treatment was included as a fixed factor (placebo, SSRI) whereas baseline severity was included both as a fixed factor and as a covariate, respectively.

In paper VI we assessed whether the symptom response profile differed between SSRIs and duloxetine when analyses were adjusted for overall antidepressant efficacy using MMRM. Separate analyses were run for each item, and the analyses included treatment outcome, represented by the sum-score of all items not serving as the dependent variable, as covariate. As the SSRI and duloxetine data come from different trials the comparison was indirect. Relative mean differences compared to placebo were computed separately for both populations (SSRI, duloxetine), and an unpaired two-sample t-test was used to assess the difference in mean differences from placebo between the two populations.

5.3 Statistical software

Offline analyses have been conducted using IBM SPSS Statistics for Mac/Windows, Version 21.0 (IBM Corp, Armonk, NY, USA) and SAS software for Windows, Version 9.4 (SAS Institute, Cary, NC, USA). For papers III and VI, remote desktop access to the Clinical Trial Data Transparency environment, SAS version 9.4, was provided by the Clinical Study Data Request website through SAS Solutions OnDemand.

Acknowledgement

First and foremost, **Elias Eriksson**, during these years I have come to greatly admire the breadth of your knowledge, your impressive talent for writing, debating and public speaking, and above all, your ability to near-instantly get to the core of any issue, no matter how abstruse. I am sincerely grateful for this time and I hope that we will continue to work together in the future. You have been, and continue to be, an inspiration.

Staffan Nilsson, while Elias knows almost all that is worth knowing about psychopharmacology, the same cannot be said about his grasp of statistics. Considering what the nuts and bolts of my PhD project have been, you can imagine my relief at having you on board. Thank you for all the time and effort that you have put into this thesis.

Hans Nissbrandt, thank you for the many spirited discussions; including, but not limited to, certain philosophical debates with my main supervisor.

Daniel Eckernäs and **Jakob Näslund**, I understand that it may come as a surprise to both of you to be referred to as a unit. And while there are indeed minor differences between the two of you – sniping prowess and command of the Russian language comes to mind – you are to me very much one and the same. Brilliant, entertaining, supportive and dependable. Getting to know you have been a pleasure and an honour.

Alexander Lisinski, we've known each other for a long time now and I feel that our relationship has evolved beyond that of mere friends … you are now also a useful asset! A sincere thank you for the hard work and many great ideas that you have contributed to this thesis.

Johan Fredrik Emilsson, I have not yet gotten my hunting license, I am, however, sure that I will get it at some point. For this, and for much other inspiration and entertainment, I am highly grateful.

To all past and present members of the Elias Eriksson group that have provided stimulating and entertaining companionship during my PhD, **Erik Studer**, **Melker Hagsäter**, **Robert Pettersson**, **Jaroslav**, **Christin**, various iterations of **highly amicable summer students**, etc. Thank you.

Daniel Hovey, you were an entertaining, friendly and highly helpful roommate for several years but then you just vanished? Curiously, the stapler that I so associate with you did not…

Martin Johnson, I think it will come in handy to have this in writing: *dance music in general, and rave in particular, is not a viable strategy against fascism*.

Camilla Fardell, whether it be coffee breaks, lunches, or quizzes, you tend to be there as a source of joy, mirth, and clever anagrams. Thank you for being you!

Which brings me to **Sittquizzarna**, and to all my other friends, who have been there for quizzes, board game nights, Overwatch evenings, a pint at the pub, etc. A heartfelt thanks to all of you for providing much needed distraction, companionship, and entertainment!

To **Lina**, **Nina**, **Daniel V**, **Daniel G**, **Jesper V**, **Jesper S**, **Yohanna**, **Johanna**, **Lars**, **Agneta**, **Gunnar**, **Mike**, **Patrik**, **Elisabet, Marie, Martina, Christer, Elin**, **Thomas, Soaleha**, **Pierre**, **Jolie**, **Britt-Marie**, **Anna-Lena**, **Rigmor**, and everyone else at the **Department of Pharmacology** who have helped and/or entertained me during the years.

Simone, you mean everything to me. Which is fitting since you have also done everything for me during these last weeks. The sincere gratitude I feel in that regard, however, pales in face of how infinitely lucky I feel for having you in my life. I love you.

Finally, my mother **Anna Karin**, my father **Detlef**, and my brothers **Johan** and **Magnus**, I trust you all know how important you are to me. I could not have asked for better friends or better role models. Thank you for your endless love, help, and support!

References

1. Paykel ES. Basic concepts of depression. *Dialogues Clin Neurosci.* 2008;**10**: 279-89.
2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders : DSM-5. 2013.
3. World Health Organization. International statistical classification of diseases and related health problems. 2016.
4. Andrade L, Caraveo-Anduaga JJ, Berglund P, et al. The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *Int J Methods Psychiatr Res.* 2003; **12**: 3-21.
5. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA.* 2003; **289**: 3095-105.
6. Hasin DS, Goodwin RD, Stinson FS, Grant BF. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Arch Gen Psychiatry.* 2005; **62**: 1097-106.
7. Hasin DS, Sarvet AL, Meyers JL, et al. Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the United States. *JAMA Psychiatry.* 2018; **75**: 336-46.
8. Waraich P, Goldner EM, Somers JM, Hsu L. Prevalence and incidence studies of mood disorders: a systematic review of the literature. *Can J Psychiatry.* 2004; **49**: 124-38.
9. Burcusa SL, Iacono WG. Risk for recurrence in depression. *Clin Psychol Rev.* 2007; **27**: 959-85.
10. Saveanu RV, Nemeroff CB. Etiology of depression: genetic and environmental factors. *Psychiatr Clin North Am.* 2012; **35**: 51-71.
11. Lohoff FW. Overview of the genetics of major depressive disorder. *Curr Psychiatry Rep.* 2010; **12**: 539-46.
12. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet.* 2018; **50**: 668-81.
13. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet.* 2007; **370**: 851-8.
14. Rutledge T, Reis VA, Linke SE, Greenberg BH, Mills PJ. Depression in heart failure a meta-analytic review of prevalence, intervention effects, and associations with clinical outcomes. *J Am Coll Cardiol.* 2006; **48**: 1527-37.
15. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. *Diabetes Care.* 2001; **24**: 1069-78.

16. Rasic DT, Belik SL, Bolton JM, Chochinov HM, Sareen J. Cancer, mental disorders, suicidal ideation and attempts in a large community sample. *Psychooncology.* 2008; **17**: 660-7.
17. Mitchell AJ, Chan M, Bhatti H, et al. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. *Lancet Oncol.* 2011; **12**: 160-74.
18. Lichtman JH, Bigger JT, Jr., Blumenthal JA, et al. Depression and coronary heart disease: recommendations for screening, referral, and treatment: a science advisory from the American Heart Association Prevention Committee of the Council on Cardiovascular Nursing, Council on Clinical Cardiology, Council on Epidemiology and Prevention, and Interdisciplinary Council on Quality of Care and Outcomes Research: endorsed by the American Psychiatric Association. *Circulation.* 2008; **118**: 1768-75.
19. Robinson RG, Jorge RE. Post-Stroke Depression: A Review. *Am J Psychiatry.* 2016; **173**: 221-31.
20. Wells KB, Sherbourne CD. Functioning and utility for current health of patients with depression or chronic medical conditions in managed, primary care practices. *Arch Gen Psychiatry.* 1999; **56**: 897-904.
21. Ferrari AJ, Charlson FJ, Norman RE, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med.* 2013; **10**: e1001547.
22. Judd LL, Akiskal HS, Zeller PJ, et al. Psychosocial disability during the long-term course of unipolar major depressive disorder. *Arch Gen Psychiatry.* 2000; **57**: 375-80.
23. Judd LL, Schettler PJ, Solomon DA, et al. Psychosocial disability and work role function compared across the long-term course of bipolar I, bipolar II and unipolar major depressive disorders. *J Affect Disord.* 2008; **108**: 49-58.
24. Harris EC, Barraclough B. Suicide as an outcome for mental disorders. A meta-analysis. *Br J Psychiatry.* 1997; **170**: 205-28.
25. Meijer A, Conradi HJ, Bos EH, Thombs BD, van Melle JP, de Jonge P. Prognostic association of depression following myocardial infarction with mortality and cardiovascular events: a meta-analysis of 25 years of research. *Gen Hosp Psychiatry.* 2011; **33**: 203-16.
26. Meijer A, Conradi HJ, Bos EH, et al. Adjusted prognostic association of depression following myocardial infarction with mortality and cardiovascular events: individual patient data meta-analysis. *Br J Psychiatry.* 2013; **203**: 90-102.
27. Pinquart M, Duberstein PR. Depression and cancer mortality: a meta-analysis. *Psychol Med.* 2010; **40**: 1797-810.
28. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet.* 2013; **382**: 1575-86.

29. Kendler KS, Engstrom EJ. Kahlbaum, Hecker, and Kraepelin and the Transition From Psychiatric Symptom Complexes to Empirical Disease Forms. *Am J Psychiatry.* 2017; **174**: 102-9.
30. Kendler KS, Munoz RA, Murphy G. The development of the Feighner criteria: a historical perspective. *Am J Psychiatry.* 2010; **167**: 134-42.
31. Wilson M. DSM-III and the transformation of American psychiatry: a history. *Am J Psychiatry.* 1993; **150**: 399-410.
32. Beck, AT. Reliability of psychiatric diagnoses: 1. A critique of systematic studies. *American Journal of Psychiatry.* 1962; **119**: 210-6.
33. Lieblich SM, Castle DJ, Pantelis C, Hopwood M, Young AH, Everall IP. High heterogeneity and low reliability in the diagnosis of major depression will impair the development of new drugs. *BJPsych Open.* 2015; **1**: e5-e7.
34. Feighner JP, Robins E, Guze SB, Woodruff RA, Jr., Winokur G, Munoz R. Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry.* 1972; **26**: 57-63.
35. American Psychiatric Association. DSM-II : Diagnostic and statistical manual of mental disorders; 1968.
36. American Psychiatric Association. DSM-III : Diagnostic and statistical Manual of Mental Disorders. Washington: D.C., APA; 1980.
37. Gelenberg AJ, Thase ME, Meyer RE, et al. The history and current state of antidepressant clinical trial design: a call to action for proof-of-concept studies. *J Clin Psychiatry.* 2008; **69**: 1513-28.
38. Regier DA, Narrow WE, Clarke DE, et al. DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses. *Am J Psychiatry.* 2013; **170**: 59-70.
39. Chmielewski M, Clark LA, Bagby RM, Watson D. Method matters: Understanding diagnostic reliability in DSM-IV and DSM-5. *J Abnorm Psychol.* 2015; **124**: 764-9.
40. Ruscio J, Ruscio AM. Informing the continuity controversy: a taxometric analysis of depression. *J Abnorm Psychol.* 2000; **109:** 473-87.
41. Hankin BL, Fraley RC, Lahey BB, Waldman ID. Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. *J Abnorm Psychol.* 2005; **114**: 96-110.
42. Ostergaard SD, Jensen SO, Bech P. The heterogeneity of the depressive syndrome: when numbers get serious. *Acta Psychiatr Scand.* 2011; **124**: 495-6.
43. Fried EI, Nesse RM. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *J Affect Disord.* 2015; **172**: 96-102.
44. Lowe B, Spitzer RL, Williams JB, Mussell M, Schellberg D, Kroenke K. Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. *Gen Hosp Psychiatry.* 2008; **30**: 191-9.

45. Hertenstein E, Feige B, Gmeiner T, et al. Insomnia as a predictor of mental disorders: A systematic review and meta-analysis. *Sleep Medicine Reviews.* 2019; **43**: 96-105.
46. Woody ML, Gibb BE. Integrating NIMH Research Domain Criteria (RDoC) into Depression Research. *Curr Opin Psychol.* 2015; **4**: 6-12.
47. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med.* 2013; **11**: 126.
48. van Praag HM. The DSM-IV (depression) classification: to be or not to be? *J Nerv Ment Dis.* 1990; **178**: 147-9.
49. Hyman SE. Can neuroscience be integrated into the DSM-V? *Nature Reviews Neuroscience.* 2007; **8**: 725.
50. Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry.* 2010; **167**: 748-51.
51. Insel TR. The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. *Am J Psychiatry.* 2014; **171**: 395-7.
52. Parker G. Antidepressants on trial: how valid is the evidence? *Br J Psychiatry.* 2009; **194**: 1-3.
53. Montgomery SA, Lecrubier Y. Is severe depression a separate indication? ECNP Consensus Meeting September 20, 1996, Amsterdam. European College of Neuropsychopharmacology. *Eur Neuropsychopharmacol.* 1999; **9**: 259-64.
54. Klerman GL, Cole JO. Clinical Pharmacology of Imipramine and Related Antidepressant Compounds. *Pharmacol Rev.* 1965; **17**: 101-41.
55. Paykel ES, Hollyman JA, Freeling P, Sedgwick P. Predictors of therapeutic benefit from amitriptyline in mild depression: a general practice placebo-controlled trial. *J Affect Disord.* 1988; **14**: 83-95.
56. Guy W. ECDEU assessment manual for psychopharmacology. Washington, D.C.: U.S. Dept. of Health, Education and Welfare, Public Health Service; 1976.
57. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960; **23**: 56-62.
58. Snaith P. What do depression rating scales measure? *Br J Psychiatry.* 1993; **163**: 293-8.
59. Snaith RP, Taylor CM. Rating scales for depression and anxiety: a current perspective. *Br J Clin Pharmacol.* 1985; **19 Suppl 1**: 17S-20S.
60. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry.* 1979; **134**: 382-9.
61. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry.* 1961; **4**: 561-71.
62. Carroll BJ, Feinberg M, Smouse PE, Rawson SG, Greden JF. The Carroll rating scale for depression. I. Development, reliability and validation. *Br J Psychiatry.* 1981; **138**: 194-200.

63. Svanborg P, Asberg M. A comparison between the Beck Depression Inventory (BDI) and the self-rating version of the Montgomery Asberg Depression Rating Scale (MADRS). *J Affect Disord.* 2001; **64**: 203-16.
64. Demyttenaere K, De Fruyt J. Getting what you ask for: on the selectivity of depression rating scales. *Psychother Psychosom.* 2003; **72**: 61-70.
65. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry.* 2004; **161**: 2163-77.
66. Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol.* 1967; **6**: 278-96.
67. Paykel ES. The clinical interview for depression. Development, reliability and validity. *J Affect Disord.* 1985; **9**: 85-96.
68. Rosenthal SH, Klerman GL. Endogenous features of depression in women. *Can Psychiatr Assoc J.* 1966; **11 Suppl**: Suppl:11-6.
69. Zitman FG, Mennen MFG, Griez E, Hooijer C. The Different Versions of the Hamilton Depression Rating Scale. In: Bech P, Coppen A, editors. The Hamilton Scales; 1990 1990//; Berlin, Heidelberg: Springer Berlin Heidelberg; 1990. p. 28-34.
70. Bech P, Allerup P, Gram LF, et al. The Hamilton depression scale. Evaluation of objectivity using logistic models. *Acta Psychiatr Scand.* 1981; **63**: 290-9.
71. Bech P, Gram LF, Dein E, Jacobsen O, Vitger J, Bolwig TG. Quantitative rating of depressive states. *Acta Psychiatr Scand.* 1975; **51**: 161-70.
72. Moncrieff J. Are antidepressants as effective as claimed? No, they are not effective at all. *Can J Psychiatry.* 2007; **52**: 96-7; discussion 102.
73. Isacsson G, Adler M. Randomized clinical trials underestimate the efficacy of antidepressants in less severe depression. *Acta Psychiatr Scand.* 2012; **125**: 453-9.
74. Williams JB. Standardizing the Hamilton Depression Rating Scale: past, present, and future. *Eur Arch Psychiatry Clin Neurosci.* 2001; **251 Suppl 2**: II6-12.
75. Fried EI. The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *J Affect Disord.* 2017; **208**: 191-7.
76. Riskind JH, Beck AT, Brown G, Steer RA. Taking the measure of anxiety and depression. Validity of the reconstructed Hamilton scales. *J Nerv Ment Dis.* 1987; **175**: 474-9.
77. Hamilton M, White JM. Clinical syndromes in depressive states. *J Ment Sci.* 1959; **105**: 985-98.
78. Spearman C. Correlation calculated from faulty data. *Br J Psychol.* 1910; **3**: 271-95.
79. Brown W. Some experimental results in the correlation of mental abilities. *Br J Psychol.* 1910; **3**: 296-322.
80. Ferguson JM. SSRI Antidepressant Medications: Adverse Effects and Tolerability. *Prim Care Companion J Clin Psychiatry.* 2001; **3**: 22-7.

81. Richelson E. Tricyclic antidepressants and histamine H1 receptors. *Mayo Clin Proc.* 1979; **54**: 669-74.
82. Ferguson JM. Fluoxetine-induced weight loss in overweight, nondepressed subjects. *Am J Psychiatry.* 1986; **143**: 1496.
83. Montgomery SA, Baldwin DS, Riley A. Antidepressant medications: a review of the evidence for drug-induced sexual dysfunction. *J Affect Disord.* 2002; **69**: 119-40.
84. Montgomery SA. The failure of placebo-controlled studies. ECNP Consensus Meeting, September 13, 1997, Vienna. European College of Neuropsychopharmacology. *Eur Neuropsychopharmacol.* 1999; **9**: 271-6.
85. Fried EI, Nesse RM. The impact of individual depressive symptoms on impairment of psychosocial functioning. *PLoS One.* 2014; **9**: e90311.
86. Zimmerman M, Martinez JA, Attiullah N, et al. Why do some depressed outpatients who are in remission according to the Hamilton Depression Rating Scale not consider themselves to be in remission? *J Clin Psychiatry.* 2012; **73**: 790-5.
87. Zimmerman M, Martinez J, Attiullah N, Friedman M, Toba C, Boerescu DA. Why do some depressed outpatients who are not in remission according to the hamilton depression rating scale nonetheless consider themselves to be in remission? *Depress Anxiety.* 2012; **29**: 891-5.
88. Klein DF, Fink M. Multiple Item Factors as Change Measures in Psychopharmacology. *Psychopharmacologia.* 1963; **4**: 43-52.
89. Faries D, Herrera J, Rayamajhi J, DeBrota D, Demitrack M, Potter WZ. The responsiveness of the Hamilton Depression Rating Scale. *J Psychiatr Res.* 2000; **34**: 3-10.
90. Mallinckrodt CH, Meyers AL, Prakash A, Faries DE, Detke MJ. Simple options for improving signal detection in antidepressant clinical trials. *Psychopharmacol Bull.* 2007; **40**: 101-14.
91. Santen G, Gomeni R, Danhof M, Della Pasqua O. Sensitivity of the individual items of the Hamilton depression rating scale to response and its consequences for the assessment of efficacy. *J Psychiatr Res.* 2008; **42**: 1000-9.
92. Turner T. Chlorpromazine: unlocking psychosis. *The BMJ.* 2007; **334 Suppl 1**: s7.
93. Selikoff IJ, Robitzek EH, Ornstein GG. Treatment of pulmonary tuberculosis with hydrazide derivatives of isonicotinic acid. *JAMA.* 1952; **150**: 973-80.
94. Lopez-Munoz F, Alamo C. Monoaminergic neurotransmission: the history of the discovery of antidepressants from 1950s until today. *Curr Pharm Des.* 2009; **15**: 1563-86.
95. Zeller EA, Barsky J, Fouts JR, Kirchheimer WF, Van Orden LS. Influence of isonicotinic acid hydrazide (INH) and 1-isonicotinyl-2-isopropyl hydrazide (IIH) on bacterial and mammalian enzymes. *Experientia.* 1952; **8**: 349-50.

96. Crane GE. Iproniazid (marsilid) phosphate, a therapeutic agent for mental disorders and debilitating diseases. *Psychiatr Res Rep Am Psychiatr Assoc.* 1957; **8**: 142-52.
97. Ayd FJ, Jr. A preliminary report on marsilid. *Am J Psychiatry.* 1957; **114**: 459.
98. Delay J, Deniker P, Buisson JF, Haim A. [Treatment of depressive states by the isonicotinic acid derivatives isoniazid & iproniazid]. *Ann Med Psychol (Paris).* 1959; **117**: 125-32.
99. Loomer HP, Saunders JC, Kline NS. A clinical and pharmacodynamic evaluation of iproniazid as a psychic energizer. *Psychiatr Res Rep Am Psychiatr Assoc.* 1957; **8**: 129-41.
100. Jacobsen E. The early history of psychotherapeutic drugs. *Psychopharmacology (Berl).* 1986; **89**: 138-44.
101. Kuhn R. The treatment of depressive states with G 22355 (imipramine hydrochloride). *Am J Psychiatry.* 1958; **115**: 459-64.
102. Axelrod J, Whitby LG, Hertting G. Effect of psychotropic drugs on the uptake of H3-norepinephrine by tissues. *Science.* 1961; **133**: 383-4.
103. Glowinski J, Axelrod J. Inhibition of Uptake of Tritiated-Noradrenaline in the Intact Rat Brain by Imipramine and Structurally Related Compounds. *Nature.* 1964; **204**: 1318-9.
104. Brodie BB, Pletscher A, Shore PA. Evidence that serotonin has a role in brain function. *Science.* 1955; **122**: 968.
105. Holzbauer M, Vogt M. Depression by reserpine of the noradrenaline concentration in the hypothalamus of the cat. *J Neurochem.* 1956; **1**: 8-11.
106. Freis ED. Mental depression in hypertensive patients treated for long periods with large doses of reserpine. *N Engl J Med.* 1954; **251**: 1006-8.
107. Achor RW, Hanson NO, Gifford RW, Jr. Hypertension treated with Rauwolfia serpentina (whole root) and with reserpine; controlled study disclosing occasional severe depression. *JAMA.* 1955; **159**: 841-5.
108. Muller JC, Pryor WW, Gibbons JE, Orgain ES. Depression and anxiety occurring during Rauwolfia therapy. *JAMA.* 1955; **159**: 836-9.
109. Harris TH. Depression induced by Rauwolfia compounds. *Am J Psychiatry.* 1957; **113**: 950.
110. Schildkraut JJ. The catecholamine hypothesis of affective disorders: a review of supporting evidence. *Am J Psychiatry.* 1965; **122**: 509-22.
111. Carlsson A, Fuxe K, Ungerstedt U. The effect of imipramine on central 5-hydroxytryptamine neurons. *J Pharm Pharmacol.* 1968; **20**: 150-1.
112. Lapin IP, Oxenkrug GF. Intensification of the central serotoninergic processes as a possible determinant of the thymoleptic effect. *Lancet.* 1969; **1**: 132-6.
113. Roth BL, Sheffler DJ, Kroeze WK. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat Rev Drug Discov.* 2004; **3**: 353-9.
114. Gillman PK. Tricyclic antidepressant pharmacology and therapeutic drug interactions updated. *Br J Pharmacol.* 2007; **151**: 737-48.

115. Preskorn SH. Therapeutic drug monitoring of tricyclic antidepressants: a means of avoiding toxicity. *Psychopharmacol Ser.* 1989; **7**: 237-43.
116. Woolf AD, Erdman AR, Nelson LS, et al. Tricyclic antidepressant poisoning: an evidence-based consensus guideline for out-of-hospital management. *Clin Toxicol (Phila).* 2007; **45**: 203-33.
117. Cassidy S, Henry J. Fatal toxicity of antidepressant drugs in overdose. *The BMJ.* 1987; **295**: 1021-4.
118. Gunnell D, Ashby D. Antidepressants and suicide: what is the balance of benefit and harm. *The BMJ* 2004; **329**: 34-8.
119. Carlsson A, Wong DT. Correction: a note on the discovery of selective serotonin reuptake inhibitors. *Life Sci.* 1997; **61**: 1203.
120. Carlsson A. A paradigm shift in brain research. *Science.* 2001; **294**: 1021-4.
121. Fagius J, Osterman PO, Siden A, Wiholm BE. Guillain-Barre syndrome following zimeldine treatment. *J Neurol Neurosurg Psychiatry.* 1985; **48**: 65-9.
122. Wenthur CJ, Bennett MR, Lindsley CW. Classics in Chemical Neuroscience: Fluoxetine (Prozac). *ACS Chemical Neuroscience.* 2014; **5**: 14-23.
123. Ludwig J, Marcotte DE. Anti-depressants, suicide, and drug regulation. *J Policy Anal Manag.* 2005; **24**: 249-72.
124. Isacsson G. Suicide prevention--a medical breakthrough? *Acta Psychiatr Scand.* 2000; **102**: 113-7.
125. Guaiana G, Barbui C, Hotopf M. Amitriptyline for depression. *Cochrane Database Syst Rev.* 2007; (3): CD004186.
126. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet.* 2018; **391**: 1357-66.
127. Montgomery SA, Bebbington P, Cowen P, et al. Guidelines for treating depressive illness with antidepressants: A statement from the British Association for Psychopharmacology. *J Psychopharmacol.* 1993; **7**(1 Suppl): 19-23.
128. Khan A, Khan S, Brown WA. Are placebo controls necessary to test new antidepressants and anxiolytics? *Int J Neuropsychopharmacol.* 2002; **5**: 193-7.
129. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *The BMJ* 2003; **326**: 1171-3.
130. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med.* 2008; **358**: 252-60.

131.Rogers SC, Clay PM. A statistical review of controlled trials of imipramine and placebo in the treatment of depressive illnesses. *Br J Psychiatry.* 1975; **127**: 599-603.
132.Brody B, Leon AC, Kocsis JH. Antidepressant clinical trials and subject recruitment: just who are symptomatic volunteers? *Am J Psychiatry.* 2011; **168**: 1245-7.
133.Posternak MA, Zimmerman M. Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials: meta-analysis. *Br J Psychiatry.* 2007; **190**: 287-92.
134.Zimmerman M, Chelminski I, Posternak MA. Exclusion criteria used in antidepressant efficacy trials: consistency across studies and representativeness of samples included. *J Nerv Ment Dis.* 2004; **192**: 87-94.
135.Moller HJ. Isn't the efficacy of antidepressants clinically relevant? A critical comment on the results of the metaanalysis by Kirsch et al. 2008. *Eur Arch Psychiatry Clin Neurosci.* 2008; **258**: 451-5.
136.Landin R, DeBrota DJ, DeVries TA, Potter WZ, Demitrack MA. The impact of restrictive entry criterion during the placebo lead-in period. *Biometrics.* 2000; **56**: 271-8.
137.Liu KS, Snavely DB, Ball WA, Lines CR, Reines SA, Potter WZ. Is bigger better for depression trials? *J Psychiatr Res.* 2008; **42**: 622-30.
138.Kobak KA, Leuchter A, DeBrota D, et al. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. *J Clin Psychopharmacol.* 2010; **30**: 193-7.
139.Mundt JC, Greist JH, Jefferson JW, et al. Is it easier to find what you are looking for if you think you know what it looks like? *J Clin Psychopharmacol.* 2007; **27**: 121-5.
140.Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med.* 2012; **367**: 1355-60.
141.Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. *The BMJ.* 2013; **346**: e8668.
142.Mallinckrodt CH, Watkin JG, Molenberghs G, Carroll RJ. Choice of the primary analysis in longitudinal clinical trials. *Pharm Stat.* 2004; **3**: 161-9.
143.Mallinckrod† CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. *Drug Inf J.* 2008; **42**: 303-19.
144.U.S. Food and Drug Administration. Levomilnacripan, statistical review. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2013/204168Orig1s000StatR.pdf. Accessed January 20, 2019.
145.Ferguson CJ. An effect size primer: A guide for clinicians and researchers. *Prof Psychol Res Pr.* 2009; **40**: 532-8.
146.Cuijpers P, Turner EH, Koole SL, van Dijke A, Smit F. What is the threshold for a clinically relevant effect? The case of major depressive disorders. *Depress Anxiety.* 2014; **31**: 374-8.

147. Turner EH, Rosenthal R. Efficacy of antidepressants. *The BMJ* 2008; **336**: 516-7.
148. Moncrieff J, Kirsch I. Empirically derived criteria cast doubt on the clinical significance of antidepressant-placebo differences. *Contemp Clin Trials.* 2015; **43**: 60-2.
149. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med.* 2008; **5**: e45.
150. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry.* 2006; **163**: 28-40.
151. Hirschfeld RM, Montgomery SA, Aguglia E, et al. Partial response and nonresponse to antidepressant therapy: current approaches and treatment options. *J Clin Psychiatry.* 2002; **63**: 826-37.
152. Liberman R. A criticism of drug therapy in psychiatry. *Arch Gen Psychiatry.* 1961; **4**: 131-6.
153. Berlim MT, Fleck MP, Shorter E. Notes on antipsychiatry. *Eur Arch Psychiatry Clin Neurosci.* 2003; **253**: 61-7.
154. Kent SA. The Globalization of Scientology: Influence, Control and Opposition in Transnational Markets. *Religion.* 1999; **29**: 147-69.
155. Citizens Commision on Human Rights. Chronology of psychiatry's role in creating the holocaust. https://www.cchr.org/documentaries/age-of-fear/creating-the-holocaust.html.Accessed January 20, 2019.
156. CBS 60 Minutes. Treating Depression: Is There A Placebo Effect? https://www.cbsnews.com/news/treating-depression-is-there-a-placebo-effect/. Accessed January 20, 2019.
157. Marcia A. The Epidemic of Mental Illness: Why? *New York Review of Books.* 2011; **58**: 20-2.
158. Begley S. Why Antidepressants Are No Better Than Placebos (Newsweek). https://www.newsweek.com/why-antidepressants-are-no-better-placebos-71111. Accessed January 20, 2019.
159. Laurance J. Antidepressants don't work - official study (The Independent). https://www.independent.co.uk/life-style/health-and-families/health-news/antidepressant-drugs-dont-work-ndash-official-study-787264.html. Accessed January 20, 2019.
160. Boseley S. Prozac, used by 40m people, does not work say scientists (The Guardian). https://www.theguardian.com/society/2008/feb/26/mentalhealth.medicalresearch. Accessed January 20, 2019.
161. Smyth C. Antidepressants do more harm than good, research says (The Times). https://www.thetimes.co.uk/article/antidepressants-do-more-harm-than-good-research-says-80p8njbcxbd. Accessed January 20, 2019.
162. Kirsch I. The emperor's new drugs: medication and placebo in the treatment of depression. *Handb Exp Pharmacol.* 2014; **225**: 291-303.

163.Kirsch I. The emperor's new drugs : exploding the antidepressant myth. New York: Basic Books; 2011.
164.Moncrieff J, Cohen D. Do antidepressants cure or create abnormal brain states? *PLoS Med.* 2006; **3**: e240.
165.Moncrieff J. The myth of the chemical cure : a critique of psychiatric drug treatment. Basingstoke, Hampshire: Palgrave Macmillan; 2009.
166.Healy D, Whitaker C. Antidepressants and suicide: risk-benefit conundrums. *J Psychiatry Neurosci.* 2003; **28**: 331-7.
167.Healy D. Let Them Eat Prozac : the Unhealthy Relationship Between the Pharmaceutical Industry and Depression (Medicine, Culture, and History): NYU Press; 2006.
168.Whitaker R. Anatomy of an epidemic : magic bullets, psychiatric drugs, and the astonishing rise of mental illness in America; 2015.
169.Gotzsche PC. Deadly psychiatry and organised denial; 2015.
170.Püras D. Report of the Special Rapporteur on the right of everyone to the enjoyment of the highest attainable standard of physical and mental health. 2017. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G17/076/04/PDF/G1707604.pdf?OpenElement. Accessed January 20, 2019.
171.Weissman MM. The epidemiology of suicide attempts, 1960 to 1971. *Arch Gen Psychiatry.* 1974; **30**: 737-46.
172.Avery D, Winokur G. Suicide, attempted suicide, and relapse rates in depression. *Arch Gen Psychiatry.* 1978; **35**: 749-53.
173.Feuerstein TJ, Jackisch R. Why do some antidepressants promote suicide? *Psychopharmacology (Berl).* 1986; **90**: 422.
174.Teicher MH, Glod C, Cole JO. Emergence of intense suicidal preoccupation during fluoxetine treatment. *Am J Psychiatry.* 1990; **147**: 207-10.
175.Dasgupta K. Additional cases of suicidal ideation associated with fluoxetine. *Am J Psychiatry.* 1990; **147**: 1570.
176.Hoover C. Additional cases of suicidal ideation associated with fluoxetine. *Am J Psychiatry.* 1990; **147**: 1571.
177.Masand P, Gupta S, Dewan M. Suicidal ideation related to fluoxetine treatment. *N Engl J Med.* 1991; **324**: 420.
178.Rothschild AJ, Locke CA. Reexposure to fluoxetine after serious suicide attempts by three patients: the role of akathisia. *J Clin Psychiatry.* 1991; **52**: 491-3.
179.Wirshing WC, Van Putten T, Rosenberg J, Marder S, Ames D, Hicks-Gray T. Fluoxetine, akathisia, and suicidality: is there a causal connection? *Arch Gen Psychiatry.* 1992; **49**: 580-1.
180.Kent SA, Manca TA. A war over mental health professionalism: Scientology versus psychiatry. *Ment Health Relig Cult.* 2014; **17**: 1-23.
181.Beasley CM, Jr., Dornseif BE, Bosomworth JC, et al. Fluoxetine and suicide: a meta-analysis of controlled trials of treatment for depression. *The BMJ* 1991; **303**: 685-92.

182.Beasley CM, Jr., Ball SG, Nilsson ME, et al. Fluoxetine and adult suicidality revisited: an updated meta-analysis using expanded data sources from placebo-controlled trials. *J Clin Psychopharmacol.* 2007; **27**: 682-6.
183.Meyer RE, Salzman C, Youngstrom EA, et al. Suicidality and risk of suicide--definition, drug safety concerns, and a necessary target for drug development: a consensus statement. *J Clin Psychiatry.* 2010; **71**: e1-e21.
184.Rouillon F, Phillips R, Serrurier D, Ansart E, Gerard MJ. [Recurrence of unipolar depression and efficacy of maprotiline]. *Encephale.* 1989; **15**: 527-34.
185.Montgomery S, Cronholm B, Asberg M, Montgomery DB. Differential effects on suicidal ideation of mianserin, maprotiline and amitriptyline. *Br J Clin Pharmacol.* 1978; **5 Suppl 1**: 77S-80S.
186.Khan A, Khan S, Kolts R, Brown WA. Suicide rates in clinical trials of SSRIs, other antidepressants, and placebo: analysis of FDA reports. *Am J Psychiatry.* 2003; **160**: 790-2.
187.Fergusson D, Doucette S, Glass KC, et al. Association between suicide attempts and selective serotonin reuptake inhibitors: systematic review of randomised controlled trials. *The BMJ.* 2005; **330**: 396.
188.Gunnell D, Saperia J, Ashby D. Selective serotonin reuptake inhibitors (SSRIs) and suicide in adults: meta-analysis of drug company data from placebo controlled, randomised controlled trials submitted to the MHRA's safety review. *The BMJ.* 2005; **330**: 385.
189.Hammad TA, Laughren TP, Racoosin JA. Suicide rates in short-term randomized controlled trials of newer antidepressants. *J Clin Psychopharmacol.* 2006; **26**: 203-7.
190.Jakobsen JC, Katakam KK, Schou A, et al. Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. *BMC Psychiatry.* 2017; **17**: 58.
191.Braun C, Bschor T, Franklin J, Baethge C. Suicides and Suicide Attempts during Long-Term Treatment with Antidepressants: A Meta-Analysis of 29 Placebo-Controlled Studies Including 6,934 Patients with Major Depressive Disorder. *Psychother Psychosom.* 2016; **85**: 171-9.
192.Hammad TA, Laughren T, Racoosin J. Suicidality in pediatric patients treated with antidepressant drugs. *Arch Gen Psychiatry.* 2006; **63**: 332-9.
193.Gibbons RD, Brown CH, Hur K, Davis J, Mann JJ. Suicidal thoughts and behavior with antidepressant treatment: reanalysis of the randomized placebo-controlled studies of fluoxetine and venlafaxine. *Arch Gen Psychiatry.* 2012; **69**: 580-7.
194.Pedersen AG. Escitalopram and suicidality in adult depression and anxiety. *Int Clin Psychopharmacol.* 2005; **20**: 139-43.
195.Acharya N, Rosen AS, Polzer JP, et al. Duloxetine: meta-analyses of suicidal behaviors and ideation in clinical trials for major depressive disorder. *J Clin Psychopharmacol.* 2006; **26**: 587-94.

196. Baldwin D, Bullock T, Montgomery D, Montgomery S. 5-HT reuptake inhibitors, tricyclic antidepressants and suicidal behaviour. *Int Clin Psychopharmacol.* 1991; **6 Suppl 3**: 49-55; discussion -6.
197. Posner K, Buchanan J, Amira L, Yershova K, Lesser A, Goldstein E. Identification and screening of suicide risk. In: Nemeroff CB, Ruiz P, Koslow SH, eds. A Concise Guide to Understanding Suicide: Epidemiology, Pathophysiology and Prevention. Cambridge: Cambridge University Press; 2014: 17-32.
198. Garlow SJ, Kinkead B, Thase ME, et al. Fluoxetine increases suicide ideation less than placebo during treatment of adults with minor depressive disorder. *J Psychiatr Res.* 2013; **47**: 1199-203.
199. Montgomery SA. Suicide and antidepressants. *Drugs.* 1992; **43 Suppl 2**: 24-30; discussion -1.
200. Stone M, Laughren T, Jones ML, et al. Risk of suicidality in clinical trials of antidepressants in adults: analysis of proprietary data submitted to US Food and Drug Administration. *The BMJ.* 2009; **339**: b2880.
201. Barbui C, Esposito E, Cipriani A. Selective serotonin reuptake inhibitors and risk of suicide: a systematic review of observational studies. *CMAJ* 2009; **180**: 291-7.
202. Miller M, Swanson SA, Azrael D, Pate V, Sturmer T. Antidepressant dose, age, and the risk of deliberate self-harm. *JAMA Intern Med.* 2014; **174**: 899-909.
203. Brent DA. Antidepressants and pediatric depression--the risk of doing nothing. *N Engl J Med*. 2004; **351**: 1598-601.
204. Newman TB. A black-box warning for antidepressants in children? *N Engl J Med.* 2004; **351**: 1595-8.
205. Brent DA, Perper JA, Moritz G, et al. Familial risk factors for adolescent suicide: a case-control study. *Acta Psychiatr Scand.* 1994; **89**: 52-8.
206. Brent DA, Perper JA, Moritz G, et al. Psychiatric risk factors for adolescent suicide: a case-control study. *J Am Acad Child Adolesc Psychiatry.* 1993; **32**: 521-9.
207. Brent DA, Perper JA, Goldstein CE, et al. Risk factors for adolescent suicide. A comparison of adolescent suicide victims with suicidal inpatients. *Arch Gen Psychiatry.* 1988; **45**: 581-8.
208. Brent DA, Baugher M, Bridge J, Chen T, Chiappetta L. Age- and sex-related risk factors for adolescent suicide. *J Am Acad Child Adolesc Psychiatry.* 1999; **38**: 1497-505.
209. Valuck RJ, Libby AM, Orton HD, Morrato EH, Allen R, Baldessarini RJ. Spillover effects on treatment of adult depression in primary care after FDA advisory on risk of pediatric suicidality with SSRIs. *Am J Psychiatry.* 2007; **164**: 1198-205.
210. Stone MB. The FDA warning on antidepressants and suicidality--why the controversy? *N Engl J Med.* 2014; **371**: 1668-71.
211. Friedman RA. Antidepressants' black-box warning--10 years later. *N Engl J Med.* 2014; **371**: 1666-8.

212.Gibbons RD, Brown CH, Hur K, et al. Early evidence on the effects of regulators' suicidality warnings on SSRI prescriptions and suicide in children and adolescents. *Am J Psychiatry.* 2007; **164**: 1356-63.
213.Lu CY, Zhang F, Lakoma MD, et al. Changes in antidepressant use by young people and suicidal behavior after FDA warnings and media coverage: quasi-experimental study. *The BMJ* 2014; **348**: g3596.
214.Isacsson G, Ahlner J. Antidepressants and the risk of suicide in young persons--prescription trends and toxicological analyses. *Acta Psychiatr Scand.* 2014; **129**: 296-302.
215.Wheeler BW, Gunnell D, Metcalfe C, Stephens P, Martin RM. The population impact on incidence of suicide and non-fatal self harm of regulatory action against the use of selective serotonin reuptake inhibitors in under 18s in the United Kingdom: ecological study. *The BMJ.* 2008; **336**: 542-5.
216.Wheeler BW, Metcalfe C, Martin RM, Gunnell D. International impacts of regulatory action to limit antidepressant prescribing on rates of suicide in young people. *Pharmacoepidemiol Drug Saf.* 2009; **18**: 579-88.
217.Baldessarini RJ, Tondo L, Strombom IM, et al. Ecological Studies of Antidepressant Treatment and Suicidal Risks. *Harv Rev Psychiatry.* 2007; **15**: 133-45.
218.Isacsson G, Holmgren A, Osby U, Ahlner J. Decrease in suicide among the individuals treated with antidepressants: a controlled study of antidepressants in suicide, Sweden 1995-2005. *Acta Psychiatr Scand.* 2009; **120**: 37-44.
219.Isacsson G, Rich CL, Jureidini J, Raven M. The increased use of antidepressants has contributed to the worldwide reduction in suicide rates. *Br J Psychiatry.* 2010; **196**: 429-33.
220.Gibbons RD, Brown CH, Hur K, Marcus SM, Bhaumik DK, Mann JJ. Relationship between antidepressants and suicide attempts: an analysis of the Veterans Health Administration data sets. *Am J Psychiatry.* 2007; **164**: 1044-9.
221.Isacsson G, Holmgren P, Druid H, Bergman U. The utilization of antidepressants--a key issue in the prevention of suicide: an analysis of 5281 suicides in Sweden during the period 1992-1994. *Acta Psychiatr Scand*. 1997; **96**: 94-100.
222.Isacsson G, Bergman U, Rich CL. Antidepressants, depression and suicide: an analysis of the San Diego study. *J Affect Disord.* 1994; **32**: 277-86.
223.Rutz W, von Knorring L, Walinder J. Frequency of suicide on Gotland after systematic postgraduate education of general practitioners. *Acta Psychiatr Scand.* 1989; **80**: 151-4.
224.Rutz W, von Knorring L, Walinder J. Long-term effects of an educational program for general practitioners given by the Swedish Committee for the Prevention and Treatment of Depression. *Acta Psychiatr Scand.* 1992; **85**: 83-8.

225. Gibbons RD, Hur K, Bhaumik DK, Mann JJ. The relationship between antidepressant medication use and rate of suicide. *Arch Gen Psychiatry.* 2005; **62**: 165-72.
226. Tiihonen J, Lonnqvist J, Wahlbeck K, Klaukka T, Tanskanen A, Haukka J. Antidepressants and the risk of suicide, attempted suicide, and overall mortality in a nationwide cohort. *Arch Gen Psychiatry.* 2006; **63**: 1358-67.
227. Mathew SJ, Charney DS. Publication bias and the efficacy of antidepressants. *Am J Psychiatry.* 2009; **166**: 140-5.
228. Ioannidis JP. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med.* 2008; **3**: 14.
229. Hieronymus F, Emilsson JF, Nilsson S, Eriksson E. Consistent superiority of selective serotonin reuptake inhibitors over placebo in reducing depressed mood in patients with major depression. *Mol Psychiatry.* 2016; **21**: 523-30.
230. U.S. Food and Drug Administration. Citalopram hydrobromide, statistical review (part 2). https://www.accessdata.fda.gov/drugsatfda_docs/nda/98/020822a_statr_P2.pdf. Accessed January 20, 2019.
231. Entsuah R, Shaffer M, Zhang J. A critical examination of the sensitivity of unidimensional subscales derived from the Hamilton Depression Rating Scale to antidepressant drug effects. *J Psychiatr Res.* 2002; **36**: 437-48.
232. Montgomery S. Is There a Role for Switching Antidepressants in Treatment-resistant Depression? In: Kasper S, Montgomery S, eds. Treatment-resistant Depression; 2013: 91-105.
233. Jakubovski E, Varigonda AL, Freemantle N, Taylor MJ, Bloch MH. Systematic Review and Meta-Analysis: Dose-Response Relationship of Selective Serotonin Reuptake Inhibitors in Major Depressive Disorder. *Am J Psychiatry.* 2016; **173**: 174-83.
234. Papakostas GI, Charles D, Fava M. Are typical starting doses of the selective serotonin reuptake inhibitors sub-optimal? A meta-analysis of randomized, double-blind, placebo-controlled, dose-finding studies in major depressive disorder. *World J Biol Psychiatry.* 2010; **11**: 300-7.
235. Meyer JH. Imaging the serotonin transporter during major depressive disorder and antidepressant treatment. *J Psychiatry Neurosci.* 2007; **32**: 86-102.
236. Meyer JH, Wilson AA, Ginovart N, et al. Occupancy of serotonin transporters by paroxetine and citalopram during treatment of depression: a [(11)C]DASB PET imaging study. *Am J Psychiatry.* 2001; **158**: 1843-9.
237. Meyer JH, Wilson AA, Sagrati S, et al. Serotonin transporter occupancy of five selective serotonin reuptake inhibitors at different doses: an [11C]DASB positron emission tomography study. *Am J Psychiatry.* 2004; **161**: 826-35.
238. Ruhe HG, Booij J, v Weert HC, et al. Evidence why paroxetine dose escalation is not effective in major depressive disorder: a randomized

controlled trial with assessment of serotonin transporter occupancy. *Neuropsychopharmacology.* 2009; **34**: 999-1010.
239. Baker CB, Tweedie R, Duval S, Woods SW. Evidence that the SSRI dose response in treating major depression should be reassessed: a meta-analysis. *Depress Anxiety.* 2003; **17**: 1-9.
240. Baker CB, Woods SW. Is there a SSRI dose response in treating major depression? The case for re-analysis of current data and for enhancing future study design. *Depress Anxiety.* 2003; **17**: 10-8.
241. Lieberman JA, Greenhouse J, Hamer RM, et al. Comparing the effects of antidepressants: consensus guidelines for evaluating quantitative reviews of antidepressant efficacy. *Neuropsychopharmacology.* 2005; **30**: 445-60.
242. McCormack JP, Allan GM, Virani AS. Is bigger better? An argument for very low starting doses. *CMAJ.* 2011; **183**: 65-9.
243. Kirsch I, Moore TJ, Scoboria A, Nicholls SS. The emperor's new drugs: An analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prevention & Treatment* 2002; **5**(1)
244. Hieronymus F, Nilsson S, Eriksson E. A mega-analysis of fixed-dose trials reveals dose-dependency and a rapid onset of action for the antidepressant effect of three selective serotonin reuptake inhibitors. *Transl Psychiatry.* 2016; **6**: e834.
245. Hieronymus F, Eriksson E. Inclusion of Flexible-Dose Trials in the Meta-Analysis of SSRI Dose-Dependency. *Am J Psychiatry.* 2016; **173**: 836.
246. Khan A, Khan SR, Walens G, Kolts R, Giller EL. Frequency of positive studies among fixed and flexible dose antidepressant clinical trials: an analysis of the food and drug administration summary basis of approval reports. *Neuropsychopharmacology.* 2003; **28**: 552-7.
247. Benkert O, Szegedi A, Wetzel H, Staab HJ, Meister W, Philipp M. Dose escalation vs. continued doses of paroxetine and maprotiline: a prospective study in depressed out-patients with inadequate treatment response. *Acta Psychiatr Scand.* 1997; **95**: 288-96.
248. Licht RW, Qvitzau S. Treatment strategies in patients with major depression not responding to first-line sertraline treatment. *Psychopharmacology.* 2002; **161**: 143-51.
249. Schweizer E, Rynn M, Mandos LA, Demartinis N, Garcia-Espana F, Rickels K. The antidepressant effect of sertraline is not enhanced by dose titration: results from an outpatient clinical trial. *Int Clin Psychopharmacol.* 2001; **16**: 137-43.
250. Kasper S, Moller HJ, Montgomery SA, Zondag E. Antidepressant efficacy in relation to item analysis and severity of depression: a placebo-controlled trial of fluvoxamine versus imipramine. *Int Clin Psychopharmacol.* 1995; **9 Suppl 4**: 3-12.
251. Naslund J, Hieronymus F, Emilsson JF, Lisinski A, Nilsson S, Eriksson E. Incidence of early anxiety aggravation in trials of selective serotonin reuptake inhibitors in depression. *Acta Psychiatr Scand.* 2017; **136**: 343-51.

252. Letemendia FJ, Harris AD. The influence of side-effects on the reporting of symptoms. *Psychopharmacologia.* 1959; **1**: 39-47.
253. Sabshin M, Ramot J. Pharmacotherapeutic evaluation and the psychiatric setting. *AMA Arch Neurol Psychiatry.* 1956; **75**: 362-70.
254. Thomson R. Side effects and placebo amplification. *Br J Psychiatry.* 1982; **140**: 64-8.
255. Rabkin JG, Markowitz JS, Stewart J, et al. How blind is blind? Assessment of patient and doctor medication guesses in a placebo-controlled trial of imipramine and phenelzine. *Psychiatry Res.* 1986; **19**: 75-86.
256. Petkova E, Quitkin FM, McGrath PJ, Stewart JW, Klein DF. A method to quantify rater bias in antidepressant trials. *Neuropsychopharmacology.* 2000; **22**: 559-65.
257. Vitiello B, Davis M, Greenhill LL, Pine DS. Blindness of clinical evaluators, parents, and children in a placebo-controlled trial of fluvoxamine. *J Child Adolesc Psychopharmacol.* 2006; **16**: 219-25.
258. Moncrieff J, Wessely S, Hardy R. Active placebos versus antidepressants for depression. *Cochrane Database Syst Rev.* 2004; (1): CD003012.
259. Quitkin FM, Rabkin JG, Gerald J, Davis JM, Klein DF. Validity of clinical trials of antidepressants. *Am J Psychiatry.* 2000; **157**: 327-37.
260. Greenberg RP, Bornstein RF, Zborowski MJ, Fisher S, Greenberg MD. A meta-analysis of fluoxetine outcome in the treatment of depression. *J Nerv Ment Dis.* 1994; **182**: 547-51.
261. Barth M, Kriston L, Klostermann S, Barbui C, Cipriani A, Linde K. Efficacy of selective serotonin reuptake inhibitors and adverse events: meta-regression and mediation analysis of placebo-controlled trials. *Br J Psychiatry.* 2016; **208**: 114-9.
262. Finney JW, Humphreys K, Kivlahan DR, Harris AH. Why health care process performance measures can have different relationships to outcomes for patients and hospitals: understanding the ecological fallacy. *Am J Public Health.* 2011; **101**: 1635-42.
263. Rutherford BR, Sneed JR, Roose SP. Does study design influence outcome?. The effects of placebo control and treatment duration in antidepressant trials. *Psychother Psychosom.* 2009; **78**: 172-81.
264. Swedish Medical Product Agency. Clinical evaluation: Nefadar, tablets 100 mg, 200 mg, 300 mg. 1993.
265. Danish University Antidepressant Group. Paroxetine: a selective serotonin reuptake inhibitor showing better tolerance, but weaker antidepressant effect than clomipramine in a controlled multicenter study. *J Affect Disord.* 1990; **18**: 289-99.
266. Danish University Antidepressant Group. Citalopram: clinical effect profile in comparison with clomipramine. A controlled multicenter study. . *Psychopharmacology (Berl).* 1986; **90**: 131-8.
267. Danish University Antidepressant Group. Moclobemide: a reversible MAO-A-inhibitor showing weaker antidepressant effect than clomipramine in a controlled multicenter study. *J Affect Disord.* 1993; **28**: 105-16.

268.Hieronymus F, Lisinski A, Nilsson S, Eriksson E. Efficacy of selective serotonin reuptake inhibitors in the absence of side effects: a mega-analysis of citalopram and paroxetine in adult depression. *Mol Psychiatry.* 2018; **23**: 1731-6.
269.Naslund J, Hieronymus F, Lisinski A, Nilsson S, Eriksson E. Effects of selective serotonin reuptake inhibitors on rating-scale-assessed suicidality in adults with depression. *Br J Psychiatry.* 2018; **212**: 148-54.
270.Perlis RH, Brown E, Baker RW, Nierenberg AA. Clinical features of bipolar depression versus major depressive disorder in large multicenter trials. *Am J Psychiatry.* 2006; **163**: 225-31.
271.Stewart JW, Quitkin FM, Liebowitz MR, McGrath PJ, Harrison WM, Klein DF. Efficacy of desipramine in depressed outpatients. Response according to research diagnosis criteria diagnoses and severity of illness. *Arch Gen Psychiatry.* 1983; **40**: 202-7.
272.Elkin I, Shea MT, Watkins JT, et al. National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. *Arch Gen Psychiatry*. 1989; **46**: 971-82; discussion 83.
273.Elkin I, Gibbons RD, Shea MT, et al. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol.* 1995; **63**: 841-7.
274.Khan A, Brodhead AE, Kolts RL, Brown WA. Severity of depressive symptoms and response to antidepressants and placebo in antidepressant trials. *J Psychiatr Res.* 2005; **39**: 145-50.
275.Montgomery SA, Kasper S. Severe depression and antidepressants: focus on a pooled analysis of placebo-controlled studies on agomelatine. *Int Clin Psychopharmacol.* 2007; **22**: 283-91.
276.Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA.* 2010; **303**: 47-53.
277.Khan A, Bhat A, Faucett J, Kolts R, Brown WA. Antidepressant-placebo differences in 16 clinical trials over 10 years at a single site: role of baseline severity. *Psychopharmacology. (Berl)* 2011; **214**: 961-5.
278.Melander H, Salmonson T, Abadie E, van Zwieten-Boot B. A regulatory Apologia--a review of placebo-controlled studies in regulatory submissions of new-generation antidepressants. *Eur Neuropsychopharmacol.* 2008; **18**: 623-7.
279.Fountoulakis KN, Veroniki AA, Siamouli M, Moller HJ. No role for initial severity on the efficacy of antidepressants: results of a multi-meta-analysis. *Ann Gen Psychiatry.* 2013; **12**: 26.
280.Eriksson E, Hieronymus F. The alleged lack of efficacy of antidepressants in non-severe depression: a myth debunked. *Acta Psychiatr Scand.* 2018; **137**: 447-9.

281. Gibbons RD, Hur K, Brown CH, Davis JM, Mann JJ. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Arch Gen Psychiatry.* 2012; **69**: 572-9.
282. Rabinowitz J, Werbeloff N, Mandel FS, Menard F, Marangell L, Kapur S. Initial depression severity and response to antidepressants v. placebo: patient-level data analysis from 34 randomised controlled trials. *Br J Psychiatry.* 2016; **209**: 427-8.
283. Furukawa TA, Maruo K, Noma H, et al. Initial severity of major depression and efficacy of new generation antidepressants: individual participant data meta-analysis. *Acta Psychiatr Scand.* 2018; **137**: 450-8.
284. Petkova E, Tarpey T, Huang L, Deng L. Interpreting meta-regression: application to recent controversies in antidepressants' efficacy. *Stat Med.* 2013; **32**: 2875-92.
285. Feighner JP, Cohn JB, Fabre LF, Jr., et al. A study comparing paroxetine placebo and imipramine in depressed patients. *J Affect Disord.* 1993; **28**: 71-9.
286. Winokur A, Gary KA, Rodner S, Rae-Red C, Fernando AT, Szuba MP. Depression, sleep physiology, and antidepressant drugs. *Depress Anxiety.* 2001; **14**: 19-28.
287. McIntyre RS, Konarski JZ, Mancini DA, et al. Measuring the severity of depression and remission in primary care: validation of the HAMD-7 scale. *CMAJ.* 2005; **173**: 1327-34.
288. Sidor MM, Macqueen GM. Antidepressants for the acute treatment of bipolar depression: a systematic review and meta-analysis. *J Clin Psychiatry.* 2011; **72**: 156-67.
289. Mallinckrodt CH, Prakash A, Houston JP, Swindle R, Detke MJ, Fava M. Differential antidepressant symptom efficacy: placebo-controlled comparisons of duloxetine and SSRIs (fluoxetine, paroxetine, escitalopram). *Neuropsychobiology.* 2007; **56**: 73-85.
290. Daly EJ, Singh JB, Fedgchin M, et al. Efficacy and Safety of Intranasal Esketamine Adjunctive to Oral Antidepressant Therapy in Treatment-Resistant Depression: A Randomized Clinical Trial. *JAMA Psychiatry.* 2018; **75**: 139-48.
291. Canuso CM, Singh JB, Fedgchin M, et al. Efficacy and Safety of Intranasal Esketamine for the Rapid Reduction of Symptoms of Depression and Suicidality in Patients at Imminent Risk for Suicide: Results of a Double-Blind, Randomized, Placebo-Controlled Study. *Am J Psychiatry.* 2018; **175**: 620-30.
292. Arnau J, Bendayan R, Blanca MJ, Bono R. Should we rely on the Kenward–Roger approximation when using linear mixed models if the groups have different distributions? *Br J Math Stat Psychol.* 2014; **67**: 408-29.