Fachbereich Erziehungswissenschaften und Psychologie der Freien Universität Berlin

Covert Heterogeneity of Major Depressive Disorder: Depression Is More Than the Sum-Score of its Symptoms

Dissertation

zur Erlangung des akademischen Grades

Doktor der Philosophie (Dr. phil.)

Doctor of Philosophy (Ph.D.)

Vorgelegt von

Dipl.-Psych. Eiko Fried

Berlin, 2013

Erstgutachterin:

Univ.-Prof. Dr. Katja Liebal (Freie Universität Berlin)

Zweitgutachterin:

Univ.-Prof. Dr. Isabella Heuser (Charité Berlin)

Tag der Disputation: 29.01.2014

In memory of Bill Zeller

1983 - 2011

Acknowledgements

Standards in my field usually lead to fairly boring acknowledgement sections that read something like this: "Eiko Fried is supported by fellowships from the Cluster of Excellence 'Languages of Emotion' (grant no. EXC302), the German Research Foundation, and the Dahlem Research School Berlin. Data collection was supported by the grant MH095109 from the National Institute of Mental Health."

In such acknowledgements, numerous people who have contributed to my research in the last three years remain unmentioned – something I would like to amend here.

First of all, I want to thank my two supervisors Prof. Dr. Katja Liebal and Prof. Dr. Isabella Heuser for great council and advice; Martin Schultze and Kerby Shedden for statistical support and patience; and Srijan Sen for the successful collaboration regarding the medical residents data.

I would also like to thank Roman Ebel and Mr Roboto, for being the very best office mates; Katie Reaves, for the papers; and Daniela Ordonez, for actually reading through all of this.

My very special thanks goes to Randolph Nesse and the NesseLab at University of Michigan – thank you all for invaluable guidance, outstanding support, and an amazing time!

Lastly, and most importantly: thank you, Mom and Dad.

Table of Contents

English Summary

Major Depressive Disorder is one of the greatest challenges of modern health. It is the leading cause of disability worldwide, highly prevalent, often recurrent, closely related to suicide, and linked to the development of life-threatening medical conditions such as diabetes and coronary heart disease. Despite decades of research, basic questions remain unresolved: genetic studies have been unable to identify loci reliably associated with depression diagnosis and treatment response, antidepressants do not work above placebo level for the majority of patients, and field trials of the recently published Diagnostic and Statistical Manual of Mental Disorders (DSM-5) show that reliability of depression diagnosis is low.

I propose that one of the main reasons for this striking lack of progress is covert heterogeneity of depression: the current diagnostic criteria lump individuals suffering from a wide range of disparate psychiatric symptoms into one undifferentiated category. Sum-scores are used instead of individual symptoms because the disease model – derived from discoveries in the field of infectious diseases at the turn of the 19th century – remains unchallenged: depression is understood to exist outside classification systems as real entity, and believed to be the common cause for its symptoms. This, in turn, makes symptoms interchangeable indicators of one underlying disease, and justifies sum-scores: symptom number, not symptom nature matters.

In this dissertation I demonstrate that individual depressive symptoms differ from each other in three important aspects. First, in a longitudinal study of 1,289 medical students undergoing the severe and chronic stressor residency, risk factors such as sex or history of depression predict increases of specific symptoms. Second, in the same sample, symptoms exhibit marked differential increases in response to severe stress. Third, in a cross-sectional study of 3,703 depressed outpatients, symptoms differ drastically in their impacts on impairment of psychosocial functioning.

Together with evidence from numerous fields of research described throughout this dissertation, these three studies illuminate that depression symptoms are more than interchangeable indicators of an underlying disease. Symptoms are distinct phenomena with particular characteristics, and the analysis of individual symptoms reveals crucial information obfuscated by sum-scores, offers important clinical utility, will substantially facilitate our understanding of depression, and lead to more efficacious prevention and intervention strategies.

Deutsche Zusammenfassung

Depression ist eine der größten gesundheitlichen Herausforderungen unserer Zeit. Die Krankheit ist weit verbreitet, oft chronisch, und häufig mit Suizid und lebensbedrohlichen Erkrankungen wie Diabetes oder koronaren Herzkrankheiten verbunden. Trotz vieler Jahrzehnte klinischer Forschung sind viele grundsätzliche Fragen ungeklärt: so wurden bisher keine Gene identifiziert, die mit Depression zusammenhängen oder Therapieerfolg vorhersagen, Ergebnisse bildgebender Verfahren wie der Magnetresonanztomographie sind relativ inkonsistent, Antidepressiva sind bei weniger als der Hälfte der Patienten wirksam, und die Reliabilität der DSM-5 Depressionsdiagnose ist niedrig.

Einen wichtigen Grund für die anhaltenden Probleme sehe ich in der verdeckten Heterogenität (*covert heterogeneity*) depressiver Symptome: die aktuelle Diagnose umfasst eine große Anzahl von Menschen, die unter unterschiedlichen und teilweise gegensätzlichen Symptomen leiden. Psychologische Forschung und klinische Praxis haben diese einzelnen Symptome in letzten Jahrzehnten nahezu vollständig ignoriert und stattdessen Summenwerte von Symptomen verwendet, weil sie stillschweigend einem über hundert Jahre alten Krankheitsmodell folgen, welches aus der Zeit der Entdeckung von Infektionskrankheiten stammt. Depression wird demnach als Krankheit verstanden, die außerhalb unserer Klassifikationssysteme als distinkte Entität existiert, und welche die gemeinsame Ursache für Depressionssymptome darstellt. Das führt wiederum dazu, dass Symptome als austauschbare Indikatoren einer latenten Störung angesehen werden, und rechtfertigt die Benutzung von Summenwerten.

In der vorliegenden Dissertation zeige ich, dass Depressionssymptome sich in drei wichtigen Aspekten voneinander unterscheiden: (1) in einer prospektiven Untersuchung von 1289 Medizinstudenten, die ein einjähriges sehr anstrengendes Praktikum unterlaufen, sagen spezifische Risikofaktoren den Anstieg ganz bestimmter Symptome vorher; (2) in der gleichen Stichprobe variieren Symptome merklich in ihrem Anstieg über die Zeit hinweg; (3) zuletzt mache ich in einer Querschnittsstudie mit 3703 depressiven Patienten deutlich, dass Symptome sich in Bezug auf ihre Auswirkungen auf psychosoziale Beeinträchtigung drastisch voneinander unterscheiden.

Zusammen mit Belegen aus anderen Studien verdeutlichen diese Untersuchungen, dass depressive Symptome nicht einfach passive oder austauschbare Indikatoren einer latenten Störung sind, sondern distinkte Prozess die einzeln studiert werden sollten. Die Analyse individueller Symptome macht hochrelevante und bisher durch Summenwerte verschleierte Informationen sichtbar und beweist hohe klinische Nützlichkeit.

1 Introduction

"How shallow were the arbitrary definitions of ordinary psychologists! [...] He began to wonder whether we could ever make psychology so absolute a science that each little spring of life would be revealed to us."
– Oscar Wilde, The Picture of Dorian Gray

1.1 Major Depressive Disorder

Major Depressive Disorder (MDD) is one of the most common psychiatric disorders, with an estimated 12-month prevalence rate of about 6.6% and a lifetime prevalence rate of about 16.2% (Kessler et al., 2003; Kessler, Chiu, Demler, Merikangas, & Walters, 2005). It is the leading cause of disability worldwide, and among the leading disorders for global disease burden (Lopez, Mathers, Ezzati, Jamison, & Murray, 2006). In the United States alone, 19 million adults suffer from a depressive illness each year, with direct and indirect costs estimated to exceed $30 billion (Lopez et al., 2006). Depression is closely related to suicide (Berman, 2009), has been linked to the development of a variety of life-threatening medical conditions like coronary heart disease (Goldston & Baillie, 2008) and diabetes (Knol et al., 2006), and more than half of all depressed patients suffer from at least one comorbid psychiatric condition (Kessler et al., 2005).

About 60% of individuals meeting criteria for MDD as defined by the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013a) report severe or very severe impairment of functioning (Kessler et al., 2003). Impairment associated with depression is long-lasting (Hays, Wells, Sherbourne, Rogers, & Spritzer, 1995) and equal or greater than impairment caused by other common, chronic medical conditions such as diabetes, hypertension, heart attack, and congestive heart failure (Mathers & Loncar, 2006; Murray & Lopez, 1996). Moreover, depression impairs functioning in various domains such as home, workplace, friends, and family (Hirschfeld et al., 2002; Judd, Paulus, Wells, & Rapaport, 1996), in many cases severely compromising the capacity for self-care and independent living.

Depression seldom occurs as solitary episode: between 50% and 75% of all individuals diagnosed with MDD experience more than one clinically significant episode in their lifetime (McClintock, Husain, Greer, & Cullum, 2010). The number of previous episodes increases relapse probability (Angst, 1999; Mueller et al., 1999) and reduces effectiveness of antidepressant medication (Kaymaz, van Os, Loonen, & Nolen, 2008). Furthermore, chronic depression is more disabling than a single MDD episode (Paradis, Reinherz, Giaconia, & Fitzmaurice, 2006), and individuals with recurrent depression show increased comorbidities on axes I through III (Katon, 2003;

Vuorilehto, Melartin, & Isometsä, 2005). Overall, this makes depression one of the most pressing health-related problems of modern living.

1.2 Open Questions

Considering the prevalence of depression, the severe suffering and impairment of individuals diagnosed with this disorder, the psychological burden for friends and families, and the great monetary costs for societies, it is not surprising that MDD has generated large amounts of research. Notwithstanding unparalleled amounts of money, time, and effort invested into answering important questions related to depression, the reach of our understanding is very limited. Disappointing results prevail especially in the domains of reliability of diagnosis, efficacy of pharmacological drugs, and the ability to predict treatment response or depression diagnosis using genetic markers.

1.2.1 Reliability of depression diagnosis

A diagnosis can be considered *reliable* if it produces similar results under consistent conditions. Hence, reliable diagnoses are of great importance for both clinical practice and research. The DSM-5 field trials published this year estimated reliability of selected DSM-5 diagnoses in large representative clinical populations (Regier et al., 2013); reliability was assessed by measuring the degree to which two clinicians independently agreed on the presence or absence of psychiatric conditions. The trials yielded a questionable inter-rater reliability of 0.28 for MDD diagnosis, with a confidence interval (CI) of 0.20-0.35; this means that clinicians much more often disagreed than agreed with each other on the diagnosis of MDD. The degree of diagnostic uncertainty was much larger for depression than for the majority of other disorders; for instance, the inter-rater reliability for borderline personality disorder was 0.54 (CI 0.43-0.66), and 0.67 (CI 0.59-0.75) for posttraumatic stress disorder (PTSD).

1.2.2 Antidepressant efficacy

A meta-analysis of four meta-analyses of efficacy trials submitted to the U.S. Food and Drug Administration (FDA) demonstrated that antidepressants were only marginally efficacious compared to placebos (Pigott, Leventhal, Alter, & Boren, 2010). Another large study of FDA trials found clinically significant differences between placebos and antidepressants only for patients "at the upper end of the very severely depressed category" (Kirsch et al., 2008, p. 0260), but pointed out that even these differences were relatively small.

The fact that placebos are only slightly less efficacious than antidepressants for the majority of patients diagnosed with MDD is even more striking when taking into account publication and

reporting biases, both of which inflate apparent antidepressant efficacy: while 94% of the antidepressants trials that were published in scientific journals found differences between antidepressants and placebos, only 51% of the trials registered with the FDA were positive (Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008). Furthermore, it is common practice to assess treatment efficacy with several screening instruments at the same time, and researchers do often not report on the primary outcome measure in case of negative results, instead presenting positive results from secondary measures (Pigott et al., 2010).

Lastly, clinical trials assess treatment response via reductions on MDD symptom scales. However, recovery from impaired functioning – a potentially more meaningful construct of actual recovery – is known to lag substantially behind symptom recovery, and impairment of functioning often persists after symptoms remit (Greer, Kurian, & Trivedi, 2010; McKnight & Kashdan, 2009; Zimmerman et al., 2008); this further contributes to overestimating the beneficial effects of antidepressants in clinical trials.

1.2.3 Lack of genetic markers predicting antidepressant response

None of over half a million common genetic markers were associated with antidepressant response in a large study with 1,790 individuals (Tansey et al., 2012). The authors concluded that the "study was large enough that it should have been possible to find common genetic variants", and that the "fact that the study failed to find such variants suggests that such variants do not exist" (p. 10).

1.2.4 Lack of genetic associations with depression diagnosis

No single locus reached genome-wide significance in a genome-wide association study (GWAS) with 34,549 subjects (Hek, Demirkan, Lahti, & Terracciano, 2013). This is in line with numerous other large studies that have failed to identify any confirmed genetic associations for MDD (Lewis et al., 2010; Shi et al., 2011; Sullivan et al., 2009; Wray et al., 2012), and contrasts with other diagnoses like schizophrenia and bipolar disorder for which several genetic associations have been replicated (Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011; Schizophrenia Psychiatric GWAS Consortium, 2011).

1.3 Covert Heterogeneity of Depression Symptoms

The authors of the large GWAS study described above concluded that "only a large sample comprising more than 50,000 subjects may be sufficiently powered to detect genes for depressive symptoms" (Hek et al., 2013, p. 667), demonstrating that solutions for the pervasive problems psychiatry has been facing are not in sight. The field is desperate to find *any* genetic associations as

long as they are significant, no matter how irrelevant these may be for clinical practice. Looking at this particular example, the size of an effect that requires a sample of 50,000 would be close to 0, and the discriminatory power of this potential association meaningless for any practical purposes (e.g., 50.5% of depressed patients have a certain biomarker and 49.5% of controls have it).

Here I want to put forward an alternative hypothesis as to why depression research has been unable to provide answers for basic questions: *covert heterogeneity of depression symptoms*. The current diagnostic approach lumps individuals suffering from a wide range of disparate psychiatric symptoms into the same, undifferentiated category, and a lack of confirmed genetic associations and reliability of diagnosis is to be expected in such a heterogeneous group. In the DSM-5, MDD is characterized by at least one major depressive episode (MDE) that leads to clinically significant distress or impairment in social, occupational, or other important areas of life. A MDE encompasses nine criterion symptoms: (1) depressed mood, (2) markedly diminished interest or pleasure, (3) increase or decrease in either weight or appetite, (4) insomnia or hypersomnia, (5) psychomotor agitation or retardation, (6) fatigue or loss of energy, (7) feelings of worthlessness or inappropriate guilt, (8) diminished ability to think or concentrate, or indecisiveness, and (9) recurrent thoughts of death or recurrent suicidal ideation. In order to meet the diagnostic criteria for MDD, an individual has to exhibit five or more symptoms, at least one of which must be either symptom (1) or (2). Two-thirds of the nine symptoms are compounds consisting of two different symptoms (2, 3, 4, 5, 7, 8), and while half of the compounds encompass related subsymptoms (2, 7, 8), the other consist of contrasting features (3, 4, 5). This leads to 1,497 unique symptom profiles that all meet DSM-5 MDD criteria (Ostergaard, Jensen, & Bech, 2011), some of which do not even share a *single* symptom: while one patient may suffer from loss of interest, insomnia, psychomotor agitation, concentration problems, and weight loss, another may report depressed mood, suicidal ideation, hypersomnia, psychomotor retardation, and weight gain. Not surprisingly, pronounced symptomatic heterogeneity of MDD diagnosis is well established (Baumeister & Parker, 2012; Lichtenberg & Belmaker, 2010), and it has even been demonstrated that symptom profiles are not stable within individuals across time (Coryell et al., 1994; Oquendo et al., 2004). Notwithstanding heterogeneity, psychiatry has largely ignored individual symptoms in the last decades and instead focused on symptom sum-scores. Despite the absence of reliable biomarkers for depression, individuals with different symptom profiles are grouped into one undifferentiated category, and patients with dissimilar problems are condensed into one common diagnosis. While covert heterogeneity may pose a general problem in psychiatry, it is especially severe for a disorder with such profound symptomatic variability like MDD.

1.4 Overview

The following chapters elucidate why ignoring covert heterogeneity has greatly hindered progress towards a better understanding of depression, and show that the study of individual depressive symptoms reveals crucial insights previously obscured.

First, the concept of essentialism and its history in modern medicine are introduced (Chapter ›2). This is important because essentialism is the reason why depression is understood as a distinct disorder and why symptom nature is ignored in favor of sum-scores. Psychiatry and clinical psychology[1] still tacitly adhere to an outdated disease model that conceptualizes mental disorders as natural kinds: diseases such as depression and schizophrenia are believed to exist outside classification systems as real entities, allowing for a reliable diagnosis similar to measles, AIDS, or ischemic heart disease. This view has led to various assumptions about mental disorders that were uncritically accepted instead of empirically tested – assumptions that are potentially wrong. Unfortunately, many everyday decisions in research and clinical practice are based on these assumptions, making them both highly problematic and highly relevant.

Chapter ›3 reviews these consequences of an essentialist understanding of mental disorders. For instance, symptoms are added up to sum-scores that presumable represent depression severity, and statistical analyses serach for associations between such scores and results of genetic or imaging analyses – in spite of profound symptomatic variability and severe problems to validate the disease category MDD. This approach is taken because depression is understood to be the common cause of depression symptoms, which in turn makes symptoms passive and interchangeable indicators of depression.

Chapter ›4 provides evidence from numerous fields of research showing that neither essentialism nor the common cause hypothesis represent good models for depression. Depressive symptoms are more than passive indicators of an underlying disease: they vary from each other in important aspects such as genetic associations, have specific direct causal links to other symptoms, and predict depression and other disorders far into the future.

The three studies that were conducted for this dissertation are presented in Chapters ›5, ›6, and ›7. They address three important open questions of covert heterogeneity: do individual depression symptoms differ from each other in risk factors and etiology (Study 1, Chapter ›5)? Do symptoms show differential increases in responses to severe stress (Study 2, Chapter ›6)? And are de-

[1] Both psychiatry and clinical psychology presuppose natural kinds and rely on the same classification systems such as the DSM. When I refer to "psychiatry" in the following sections, I do so for the sake of brevity and legibility; problems reviewed below pertain to clinical psychology as well.

pressive symptoms differentially associated with impairment of psychosocial functioning (Study 3, Chapter ›7)?

In Chapter ›8, a new symptom-based approach to understanding and investigating depression is introduced. This model accounts for differences between symptoms on the one hand, and on the other hand allows symptoms to be directly and causally related to each other. Both benefits and limitations of the model will be discussed.

Lastly, Chapter ›0 summarizes implications of the dissertation for future depression research and explores how well problems of and solutions to covert heterogeneity generalize to other psychiatric conditions.

2 Essentialism – The Heart of the Problem

"The first step in most of the sciences is purely classificatory".
– William James, 1890, p. 646

"The classification of mental illnesses was borne out of the necessity to provide a common language for descriptive purposes and delivery of care […] although a biological basis was not a factor in the initial classification, the system has fostered a general attitude that psychiatric disorders are biologically distinct."
– Sibille & French, 2013, p. 1

2.1 The Historical Roots of Essentialism in Psychiatry

2.1.1 Infectious diseases

In the year 1890, the German microbiologist Robert Koch published his postulates that established the causative role of bacteria in the etiology of infectious diseases such as anthrax, cholera, and tuberculosis. Organisms associated with other diseases, like the syphilis bacterium *Treponema pallidum*, were discovered shortly thereafter, and medicine consolidated its understanding of medical disorders as *natural kinds* (Boyd, 1999; Kendler, Zachar, & Craver, 2010; Zachar & Kendler, 2007; Zachar, 2002). This perspective envisions diseases as unchanging and ahistorical entities that exist outside of classification systems. Diseases have sharp boundaries between each other, discrete and deterministic causes, and are defined by a specific set of properties (e.g., symptoms and duration) that are both necessary and sufficient for a diagnosis. This particular way of classification is often referred to as *essentialism* (Kendler et al., 2010; Wilson, Barker, & Brigandt, 2007), and an essence in this sense can be defined as "some kind of underlying, intrinsic property, something that lies within kind members, making them the kind of thing that they are" (Wilson et al., 2007, p. 3)[2]. All members of a kind have certain intrinsic properties, and identifying these properties allows for a reliable classification.

Chemical elements provide good examples for natural kinds: gold, for instance, has the atomic number 79, and everything with this atomic number is gold. The internal structure itself defines kind membership, not a man-made classification system. Measles represents an exemplary infectious disease readily classifiable within this framework: it is an infection of the respiratory system caused by a specific RNA virus, and accompanied by various symptoms like red eyes, fever, and a generalized rash. Moreover, many individuals suffering from measles exhibit a pathogno-

[2] Essentialism in this sense is often referred to as *kind essentialism*, and these two terms as well as the notion of natural kinds are used interchangeably in this dissertation.

monic symptom (*sine qua non*) – Koplik's spots inside the patient's mouth – that allows for a diagnosis beyond any reasonable doubt. Similar to gold, intrinsic properties reliably define what measles is.

Robert Koch won the Nobel Prize in 1905 for his groundbreaking discoveries pertaining to the role of causative agents in the development of tuberculosis, and it was soon established that infectious diseases are distinct entities that differ from each other in etiology and symptom presentation. Tuberculosis, for instance, is caused by specific mycobacteria (that differ from bacteria causing other infectious diseases), and an infection with these mycobacteria subsequently leads to symptoms such as chronic cough, fever, and night sweats (once again, these symptoms differ from symptoms of other infections). This disease model of natural kinds has been considered the most important idea in the history of nosology (Hyland, 2011), and has been crucial in the development of successful treatments. Treating tuberculosis with antibiotics is only sensible after one has realized four things: that tuberculosis is caused by specific bacterial agents; that the underlying disease causes particular symptoms; that these symptoms indicate the presence of the latent disorder; and that antibiotics are a successful in the treatment of bacterial infections, ultimately leading to the remission of tuberculosis and (as a consequence) of its symptoms.

Overall, this disease model encompasses two major assumptions: that disorders are distinct entities, and that the underlying disorder causes symptoms that can therefore be used to indicate the disorder. The following sections elaborate on the first point – essentialism –, whereas causation is the focus of Chapter ›3.

2.1.2 General paresis

The idea of diseases as natural kinds with discrete causes also worked well for the first psychiatric disease entity identified: general paresis, known at that time as *general paralysis of the insane* or *progressive paralysis*. General paresis is a neuropsychiatric syndrome present in individuals suffering from late-stage syphilis, and was considered a mental disorder upon discovery due to various psychotic symptoms. While the disease was described as early as 1822, it took roughly a century until all doubts about its causes could be laid to rest when syphilitic bacteria were identified in the brains of paretics shortly after the turn of the 19th century. The discovery was a crucial moment in the history of psychiatry because the disease model of infectious diseases was applied to and subsequently adopted for psychiatric diagnoses. This application, however, had far-reaching consequences, and two German psychiatrists critically discussed general paresis as an exemplary psychiatric condition. In 1912, Alfred Roche wrote:

> "The main example of a happy final definition of disease conditions, which in all directions constantly prove to belong together, has been progressive paralysis. The success achieved here has perhaps been a misfortune in its side effects because it nourished the illusion that something similar might soon be repeated." (Sass, 2007, p. 139)

Over 50 years later, Kurt Schneider concluded in 1959:

> "General paralysis […] became the model for forming disease entities. It was thought it would continue thus, it was hoped that with time more and more such disease entities would emerge from the multifarious conditions of the mentally ill. In fact, however, this did not happen." (Sass, 2007, p. 428)

2.2 Essentialism in Modern Psychiatry

Today, a century after the discovery of general paresis, the disease model described above is one of the most important premises for the classification of mental disorders and deeply entrenched in clinical practice and research. Disorders like depression or schizophrenia are understood to be natural kinds (similar to infectious diseases such as syphilis or tuberculosis) with essences that fundamentally define them and at the same time separate them from other disease entities. This should make it possible to reliably distinguish healthy individuals from subjects with mental disorders, and patients with one specific mental disorder from patients with another one. The widely cited paper "Depressive Disorders: Towards a Unified Hypothesis" by Akiskal and McKinney (1973) in which the authors suggested one final common pathway for *the* depressive syndrome is a good example for kind essentialism in psychiatry. This view also explains the quest for biomarkers: since disorders exist as distinct entities, are predisposed by particular genes, and manifest themselves in brain abnormalities, both the vulnerability genes and brain dysfunctions can be discovered.

Unfortunately, a Robert Koch for psychiatric conditions is not in sight: despite all the efforts, major discoveries validating psychiatric disease categories are absent. For this reason, critical voices have surfaced calling psychiatry a "semi-science" (e.g., Brooks, 2013). The president of the American Psychiatric Association, Jeffrey Lieberman, recently responded to such criticism and stated that progress "has been largely limited by technology" (Lieberman, 2013); the human genome and brain are highly complex, and identifying disturbed brain areas, dysfunctional neurotransmitter systems, and genes involved in the development and retention of mental disorders is a very difficult matter – but ultimately a matter of time and technology.

Undoubtedly, it is crucial to continue trying to elucidate the biological roots of mental disorders: new statistical approaches, larger datasets, and more powerful computational techniques

may enable us to uncover previously hidden complex effects. Additionally, the field of epigenetics – the study of gene-environment interactions – promises substantial contributions to understanding mental disorders. It is also well established that most mental disorders are at least moderately heritable (Zuk, Hechter, Sunyaev, & Lander, 2012), justifying further investments into the discovery of biomarkers. Nevertheless, it is important to note that psychiatry nearly unanimously identifies the scapegoat for the lack of progress in technological areas, while the *disease model itself* – our understanding of mental disorders – remains largely unquestioned. The next sections review general problems of essentialism as well as problems pertaining to the disease model of natural kinds.

2.3 General Problems of Essentialism

In recent years, critical voices have suggested that classification practices relying on the assumption of natural kinds are problematic. Zachar (2002) returns to chemical elements in his example of hydrogen: while all three hydrogen isotopes (protium, deuterium, and tritium) have one proton in their nuclei, they have variable numbers of neutrons that determine the specific isotope characteristics. We have decided that the number of protons should be the defining characteristic for elements; likewise, we could have chosen the number of neutrons and would have obtained a different classification system. Hydrogen exemplifies that internal criteria are often not self-sufficient for purposes of classification, and external considerations play a role – a notion that is incompatible with the hypothesis of distinct natural kinds for which defining properties are inherent features of things.

Similar objections have been raised for another well-known domain of classification: biological species. Commonly, species are thought to be distinct entities that are clearly demarcated, and that inherent properties define species membership. However, several authors have argued that the idea of distinct species is pre-Darwinian, because species are populations of animals with heterogeneous genomes (Ereshefsky, 2007; Wilson et al., 2007). Variation and heterogeneity within a species is not a deviation from the true essence of a biological kind, but part of what it is to be a member of those kinds; while species may be useful categories, they are an idealization of nature.

A third example, this time from the field of psychology, is the concept of basic or primary emotions such as fear, anger, or disgust. Such emotions are believed to exist as clearly demarcated and distinct entities by lay people (Lindquist, Gendron, Oosterwijk, & Barrett, 2013) as well as prominent researchers (e.g., Ekman & Cordaro, 2011; Panksepp, 2004); Ekman and Cordaro, for instance, argue that "emotions are discrete [and] can be distinguished fundamentally from one another" (Ekman & Cordaro, 2011, p. 364). Basic emotions are understood to have metaphysical

essences that define them and at the same time separate them from each other. However, William James already noted in 1890 that "surely there is no definite affection of 'anger' in an 'entitative' sense" (James, 1950, p. 206), and various researchers have since argued against the idea of basic emotions as natural kinds (e.g., Lindquist et al., 2013; Zachar & Bartlett, 2002; Zachar, 2002).

Studies have shown that an implicit essentialist worldview develops early in human cognition (Gelman, 2009) and applies to numerous domains of classification (Haslam, Rothschild, & Ernst, 2000; Prentice & Miller, 2007), not just chemical elements, species, and emotions; there seems to be a natural human tendency to essentialize. This phenomenon that has been termed *essentialist bias* in the psychological literature (Zachar & Bartlett, 2002) because the representation of such categories reflects perception and not necessarily reality.

2.4 Problems of Essentialism in the Realm of Mental Disorders

Kind essentialism may not only be ill-suited to describe chemical elements or biological species; it may also provide a problematic foundation for the classification of mental disorders, and several objections have been raised against the hypothesis that psychiatric diagnoses represent natural kinds (see Kendler et al., 2010; Zachar & Kendler, 2007; Zachar, 2002).

First, biological systems are highly interdependent: genes express proteins that work in cells that ultimately shape behavior – and at most levels, regulatory feedback mechanisms with the environment exist. Declaring one of these processes to be a fundamental part of the essence of a mental disorder is arbitrary and ignores the complex nature and dynamic causality of biological systems (Kendler & Baker, 2007; Zachar, 2002).

Second, approximately 45% of patients meeting the diagnostic criteria for a mental disorder receive at least one additional diagnosis (Kessler et al., 2005; Mineka, Watson, & Clark, 1998); this is hard to reconcile with the hypothesis of psychiatric conditions as distinct natural kinds.

Third, despite extraordinary monetary investments and studies of large population samples, the field has failed to identify intrinsic biological properties of mental disorders. In 1980, the DSM-III (American Psychiatric Association, 1980) preamble predicted that biomarkers associated with most diagnoses would be identified by the time the DSM-IV (American Psychiatric Association, 1994) appeared; 33 years and two DSM versions later, and with the exception of some neurological disorders, not one biological test for mental disorders was ready for inclusion in the criteria sets for the DSM-5, and not a single psychiatric diagnosis can be validated by laboratory or imaging biomarkers (Kapur, Phillips, & Insel, 2012; Nesse & Stein, 2012). Additionally, although many mental disorders are moderately heritable, effects for specific genes are small at best, and often not

specific to one diagnosis (Kendler, 2005; Purcell et al., 2009; Shi et al., 2009). This has led several authors to argue that the idea of "a gene for mental disorder x" may be inappropriate for psychiatric disorders in general (Kendler, 2005; Sibille & French, 2013).

Here I suggest that the tacit adherence to essentialism is at the very heart of many of the unsolved problems psychiatry has been facing (see Chapter ›1.2), especially in the domain of depression research. The difficulty is not so much that all clinicians and researchers actively hold an essentialist perspective – in fact, there are outspoken opponents of the disease model of natural kinds in psychiatry (e.g., Kendler et al., 2010; Zachar & Kendler, 2007; Zachar, 2002). The problem is that the notion of natural kinds is deeply rooted in the history of medicine and represents a highly plausible world-view that comes to us intuitively; as a result, numerous clinical and research practices are based on essentialist assumptions, and the way we measure, diagnose, and treat depression all presuppose that mental disorders are natural kinds. Specifically, essentialism has led to four problematic, unquestioned, and closely interconnected assumptions that are discussed in the next chapter.

3 The Common Cause Hypothesis of Depression

> *"Finding the common cause was therefore a major goal in research, with serotonin shortage being the most likely candidate. However, treatment with anti-depressants […] turned out to be beneficial for only some people, thereby ruling out serotonin as the common cause of depression symptoms […]. No other plausible common causes have ever been found, in our opinion due to the fact that there simply is no common cause that explains the entirety of depression symptoms."*
> *– Cramer et al., 2010, p.141*

Upon discovery of causative agents for infectious diseases, these disorders were understood as distinct entities. Over the years, this disease model was tacitly adopted for mental disorders, and has had two major consequences. First, mental disorders today are still assumed to be distinct entities, as discussed in Chapter ›2. Second, psychopathological symptoms are generally understood to be consequences of underlying disorders, similar to infectious diseases. This second presupposition is focus of the present chapter and referred to as *common cause hypothesis* (see Cramer, Borsboom, Aggen, & Kendler, 2011; Cramer, Waldorp, van der Maas, & Borsboom, 2010; Pearl, 2000; Schmittmann et al., 2013). The model is deeply engrained in everyday psychiatry; specifically, there are four interconnected assumptions within the common cause framework:

1. Depressive symptoms are passive indicators that are caused by the underlying depressive disorder (Chapter ›3.1).
2. Since symptoms indicate the same latent disease, they can be treated as diagnostically interchangeable in the DSM and depression screening instruments (Chapter ›3.2).
3. Depression symptoms are uncorrelated with each other once the underlying disorder is controlled for – they are consequences of the disorder, and these consequences are only correlated because of one common cause (Chapter ›3.3).
4. Symptoms are added up to sum-scores that reflect depression severity, and thresholds reliably distinguish healthy subjects from those with mental disorders (Chapter ›3.4).

3.1 Reflective Latent Variable Models

A latent variable is something that cannot be observed directly. Instead of assessing latent variables like intelligence, we measure indicators of intelligence – e.g., tests – that provide information about the latent. Statisticians differentiate between *formative* and *reflective* latent variable models. In formative models, the latent variable is determined by its measures (Bollen & Lennox, 1991; Schmittmann et al., 2013). An example is socioeconomic status (SES), which is defined by a func-

tion of observable indicators like income, education, job, and neighborhood: if a person's income increases, so does the SES, and changes in SES indicators thus lead to changes in the latent.

In reflective models, the causation goes the other way: changes in the a latent variable lead to changes in its indicators (Bollen, 1989; Bollen & Lennox, 1991; Schmittmann et al., 2013). For instance, the latent personality trait extraversion influences how likely it is that someone enjoys talking to strangers or attends social events. Extraversion is thus the common cause of its indicators, and can be assessed using these (and other) observable items[3].

In medicine, the disease model of natural kinds, the common cause hypothesis, and reflective latent variable models are closely entangled, and derived from discoveries in the field of infectious diseases: syphilis bacteria cause general paresis, general paresis causes the symptoms of general paresis, and treatment of the underlying problem will cure general paresis, resulting in the remission of symptoms. Medical disorders in general are conceptualized within this reflective framework: when a patient complains about polyuria (frequent urination), polydipsia (increased thirst), and polyphagia (increased hunger), a doctor would conclude that diabetes is the most probably common cause for the symptoms, and the diagnosis can be substantiated via biomarkers such as high blood sugar.

In psychiatry, symptom checklists and screening instruments are used to assess disorders: because depression is understood to be the common cause of its symptoms, we query individuals about depressive symptoms to investigate the presence or absence of the underlying (and not directly observable) disease entity. This perspective is perhaps most obvious in graphical representations of psychopathological latent variable models in which arrows lead from depression *to* the depressive symptoms, describing a clear direction of causation (Figure 3-1).

[3] For readers especially interested in latent variable models and causation: note that *reactive* indicators have been suggested recently that act both as cause and effect of an underlying latent variable (see Hayduk, Robinson, & Perks, 2007).

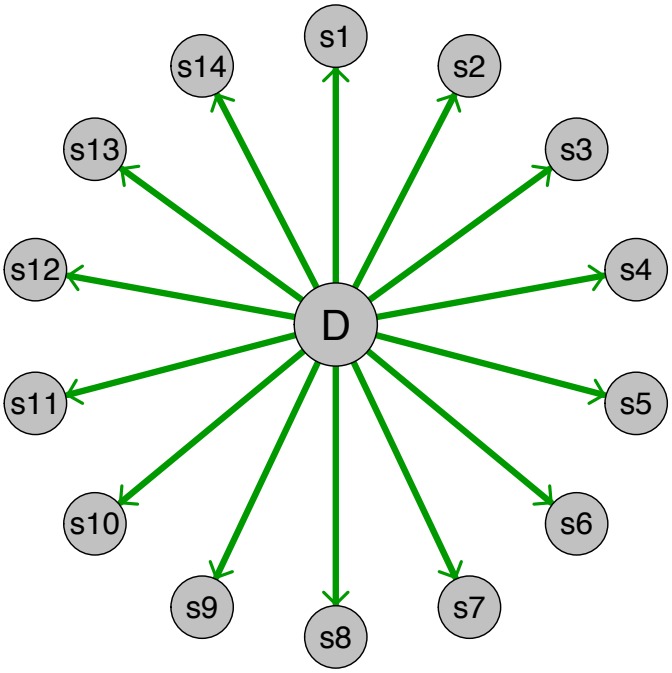

Figure 3-1. *Visualization of a reflective latent variable model.* D indicates the latent disorder depression that is the common cause of the observable symptoms s1-s14. The plot was constructed using the R-package QGRAPH (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012).

3.2 Assumption of Symptom Equivalence

Because depression is viewed as the common cause for diverse depressive symptoms such as sad mood, loss of interest, insomnia, and psychomotor agitation, symptoms are *passive* indicators of the same latent disease. The more symptoms are present, the greater the likelihood that a person actually has the latent condition, irrespective of the *particular* symptoms reported: symptom number, not symptom nature matters. Reflective latent variable models render all symptoms equally central to the disorder, and symptoms become diagnostically equivalent and interchangeable; this assumption of symptom equivalence is the main reason why lumping individuals suffering from numerous different symptoms into one diagnostic category is not considered problematic (Bollen & Lennox, 1991; Cramer et al., 2010; Lux & Kendler, 2010)[4].

[4] As discussed in Chapter ›1.3, the DSM encompasses two MDD core symptoms of which at least one has to be present for a diagnosis of depression. While this hierarchical structure implies that symptoms are not completely interchangeable, clinical screening instruments do not account for this distinction, and neither do latent variable

3.3 Assumption of Local Independence

A statistical phenomenon called *local independence* is another assumption within the common cause framework. Since depressive symptoms are a corollary of an underlying disorder, they are not correlated with each other beyond this common cause that accounts for symptom covariance (Holland & Rosenbaum, 1986; Pearl, 2000; Schmittmann et al., 2013). In other words, correlations between symptoms are *spurious* because observed item responses are statistically independent once the latent variable is controlled for.

For example, a statistical analysis might reveal that correlation between the number of firemen present at a fire and the damage caused by the fire is highly significant. Whereas a naïve observer may conclude that these variables are directly (and possibly causally) related, the correlation is spurious and seizes to exist once the size of the fire is controlled for: a massive fire will lead to a large number of firemen present as well as substantial damage caused.

For depressive symptoms, local independence seems problematic; if a person suffers from insomnia (e.g., due to chronic pain), would it not be possible that fatigue, concentration problems, weight loss, and other depressive symptoms develop in causal response to insomnia, and not as a corollary of the latent disorder?

3.4 Sum-Scores

When a clinician or researcher wants to determine whether a person is depressed or not, the DSM criterion symptoms or clinical screening instruments such as the Beck Depression Inventory (BDI; Beck et al., 1988) or the Hamilton Rating Scale for Depression (HAM-D; Hamilton, 1960) are employed. Symptoms are assessed, added up to a total score that presumably reflects depression severity, and thresholds distinguish healthy individuals from patients. This information is subsequently used to guide numerous important decisions: for instance, whether treatment is indicated, or whether an individual takes part in a study as control subject or case.

Statistically, in case all depressive symptoms do measure the same latent variable, an unweighted sum-score is either sufficient to describe the latent, or will describe it reasonably well (see Cramer et al., 2010). This is closely connected to the assumption of symptom equivalence: sum-scores can only meaningfully represent severity of the latent disorder if all summands are similar things (we are adding together apples and apples, not apples and oranges). If, however, symptoms

models: symptoms are used as equally central indicators of the same latent disorder, and thus interchangeable for all practical purposes.

are distinct phenomena and not passive indicators of one latent condition, the current approach is problematic: individuals with dissimilar symptoms will end up in the same "depressed" group in research studies, and it may not be surprising that the quest for associated genes, brain correlates, and efficacious treatment has been fairly unsuccessful, seeing that the variables that we try to predict – binary indicators (depressed vs. healthy) or symptom sum-scores – obfuscate crucial information about the nature of a patient's depression.

The following chapter describes empirical findings that are hard to reconcile with the two cornerstones of the disease model outlined above: that depression is a distinct entity, and that the underlying disorder MDD is the common cause for its symptoms.

4 Depression: Disease Model and Reality

> *"A conditionalized realism about psychiatric disorders […] is so useful because it forces us to revise our theories when the world tells us that they are wrong."*
> *– Kendler, Zachar & Craver, 2010, p. 1149*

The tacit application of an essentialist disease model to psychiatry has led to a variety of premises that seem questionable for depression, considering the pronounced variability and complex nature of MDD symptoms. These assumptions – such as the common cause hypothesis, symptoms as passive and interchangeable indicators, local independence, and sum-scores that supposedly represent MDD severity – are generally presumed to be correct instead of critically examined. In addition to more general problems such as high comorbidity rates and lack of biomarkers outlined above, the present chapter reviews research highlighting numerous problems with psychiatry's understanding of depression. Since essentialism (see Chapter ›2) and assumptions within the common cause framework (see Chapter ›3) are closely intertwined and evidence often contradictory of several assumptions at the same time, the following sections are structured into important discoveries instead of single assumptions.

4.1 Genetic Heterogeneity

Overall, three reports have documented genetic heterogeneity within the disease category MDD. Jang et al. (2004) showed that 14 depression symptoms differed from each other in their degree of heritability (h^2 range: 0 - 35%), with somatic symptoms (e.g., loss of appetite) generally showing higher heritability coefficients than affective ones (e.g., tearfulness). A second study (Myung et al., 2012) revealed that individual MDD symptoms are differentially associated with specific gene polymorphisms; for instance, the authors identified a significant association between middle insomnia (waking up during the night and having trouble to fall asleep again) and the GGCCGGGC haplotype in the first haplotype block of TPH1. In addition, a recent report of 7,500 twins identified three genetic factors that exhibited pronounced differential associations with specific MDD symptoms, concluding that the DSM symptomatic criteria for depression do not reflect a single underlying genetic process (Kendler, Aggen, & Neale, 2013). While the particular results reported above need to be replicated, especially for the two smaller studies that had less than 500 participants each (Jang et al., 2004; Myung et al., 2012), the overall findings are difficult to reconcile with the notion of depression as distinct disorder as well as the assumption of symptom equivalence. Results raise doubts about using sum-scores or the binary distinction depressed vs. healthy to investigate genetic

associations, because such analyses can only capture the shared genetic variance of all symptoms, and it is unlikely that disparate symptoms such as loss of interest and psychomotor problems share substantial proportions of genetic variance.

4.2 Lack of a Zone of Rarity

Another difficulty for the disease model of natural kinds is the lack of a *zone of rarity* between depressed and healthy individuals, as documented by a large number of taxometric analyses (e.g., Aggen, Neale, & Kendler, 2005; Kendell & Jablensky, 2003; Ruscio, Zimmerman, McGlinchey, Chelminski, & Young, 2007; Slade & Andrews, 2005). In contrast to some medical conditions such as measles or syphilis, there are no two distinct populations 'healthy' and 'depressed'. Instead, there is a continuum between healthy on the one hand and severe clinical depression on the other, and where to exactly draw the boundary is somewhat arbitrary. This is supported by evidence that sub-threshold depression is often clinically significant (Pincus, Davis, & McQueen, 1999); individuals that do not fully meet the DSM criteria regularly exhibit depression-like levels of functional impairment, suffer from psychiatric and physical comorbidities, and have an increased risk of future depressive episodes (Gotlib, Lewinsohn, & Seeley, 1995; Kendler & Gardner, 1998; Luyten, Blatt, Van Houdenhove, & Corveleyn, 2006; A. Solomon, Haaga, & Arnow, 2001). Moreover, substantial familial associations between sub-threshold and full-threshold depression have been documented (Lewinsohn, Klein, Durbin, Seeley, & Rohde, 2003). The traditional essentialist view of depression in which a person either does or does not have the disease is problematic. This classification difficulty is similar to determining a group of intellectually gifted individuals: intelligence is on a continuous scale, and the question whether to use one, two, or three standard deviations *(SD)* above the mean as threshold to demarcate highly gifted people is arbitrary and does not represent a naturally existing boundary.

Depressive symptoms themselves have been argued to be on one dimension with non-pathological phenomena (Persons, 1986). Except for suicidal ideation that likely represents a clinical symptom even in its weakest form, occasional sleep problems or sadness are perfectly normal and do not indicate the presence of a mental disorder. A dimensional approach to MDD – based on an understanding that depressive symptoms themselves are dimensional – may offer an important perspective that could help the field to overcome a simplified dichotomous thinking. As Bannister (1968) pointed out for a different diagnosis, there is a tendency to "erect psychologies of schizophrenics as if they were a logically distinct species" (p. 183); the depressive continuum runs

from none to nine DSM symptoms, and studies searching for a distinct "depression species" (e.g., in terms of a specific genotype) will likely not lead to relevant insights.

While symptom sum-scores bring various problems with them (see Chapter ›3.4), a binary approach is even more questionable: without distinct healthy and depressed populations, both groups lack validity. This further complicates the situation of covert heterogeneity: the diagnostic category MDD not only is heterogeneous in regard to the nature of symptoms reported, but also the number of symptoms, and a threshold distinguishing healthy from depressed subjects may fail to detect individuals that suffer from high levels of impairment although they do not meet the diagnostic criteria, and may falsely include others despite the presence of five or more criteria.

4.3 Depressive Symptoms and Antidepressant Efficacy

Many patients diagnosed with MDD are treated with antidepressants, e.g., selective serotonin reuptake inhibitors (SSRIs). This approach is in line with the disease model of natural kinds and the common cause framework: depression is a distinct and specific brain malfunction and manifests itself via particular symptoms. Drugs target these brain imbalances, and once the common cause of the symptoms is resolved, symptoms will disappear. Similar to other medical conditions, treating particular symptoms may relieve the patient's suffering, but is technically not necessary to ending the disease itself; the symptoms will disappear eventually once the underlying disease has been cured.

However, studies suggest that the presence of individual symptoms may have large impacts on therapy success. Of these symptoms, sleep problems – that are reported by the majority of individuals diagnosed with MDD (Benca & Peterson, 2008; Peterson & Benca, 2008) – have received the greatest attention so far. Sleep problems reduce the efficacy of depression treatment (Dew et al., 1997), patients with persistent insomnia are more than twice as likely to remain depressed (Pigeon et al., 2008), and insomnia has been shown to become chronic despite successful resolution of depressive symptoms (Mouchabac, Ferreri, Cabanac, & Bitton, 2003; Reynolds et al., 1997; Thase et al., 2002; Van Londen, Molenaar, Goekoop, Zwinderman, & Rooijmans, 1998). Moreover, and most importantly, specifically targeting sleep problems increases overall depression improvement (Lichstein, Wilson, & Johnson, 2000; Rybarczyk, Lopez, Benson, Alsten, & Stepanski, 2002). These findings are not possible within the common cause framework: treating a symptom directly cannot not diminish the severity of the latent disorder, the same way manually changing the number shown on the temperature display of a thermometer to a lower value will not reduce

the fever of a person; causality does not work that way in models in which an observable variable is a consequence of an underlying trait.

In addition to therapy studies on sleep problems, there is evidence that the presence of non-MDD symptoms may moderate treatment efficacy. Both irritability and anger have been shown to be highly prevalent clinical markers of a more severe, chronic, and complex depressive illness (Judd, Schettler, Coryell, Akiskal, & Fiedorowicz, 2013), and anxiety symptoms in depressed patients both reduce remission rates and prolong remission (Fava et al., 2008). Moreover, Uher et al. (2012) demonstrated that the symptoms 'loss of interest', 'diminished activity', and 'inability to make decisions' predicted poor antidepressant response.

This contradicts core assumptions of the disease model, and underlines the importance of utilizing symptom information to inform clinicians about the best treatment options. Ignoring specific symptoms by treating them as interchangeable indicators of a latent disease likely leads to decreased treatment efficacy and increased relapse rates. Imagine a doctor is faced with 35 patients, each reporting a unique set of three of the following seven symptoms: running nose, headaches, aching limbs, fever, coughing, sore throat, and itching eyes. A sum-score of three symptoms groups all patients into one category and leaves the doctor in the dark about whether antibiotic treatment is indicated (bacterial vs. viral infection), whether a patient has a common cold or the flu, whether antihistamines would be the correct response to alleviate symptoms of a cat allergy – or whether the patient was crying before he came to see the doctor and is in desperate need of a hug. Treating all these patients with the same medication would very likely not work significantly above placebo level – although particular forms of treatment most certainly work for individuals with *specific* symptoms.

Numerous large studies have documented surprisingly low antidepressant efficacy (e.g., Kirsch et al., 2008; Pigott et al., 2010), and clinicians are aware of the fact that some antidepressants have sedative effects, while others lead to improvement of motivation and energy[5]. Notwithstanding these facts, individual symptoms are currently not examined in treatment efficacy studies, and acknowledging and investigating the complex nature of depression is an important step to developing targeted treatments for subsets of patients.

[5] In the general treatment literature, this has been referred to as a problem of *latent heterogeneity* (Pearl, 2012), and symptom-based drug trials may be necessary to elucidate effects hidden by sum-scores (for methodological considerations, see Pearl, 2012; Xie, Brand, & Jann, 2012).

4.4 Symptoms as Side Effects of Treatment

Antidepressants are known to cause significant side effects, with prevalence rates of up to 27% in clinical trials (Trindade, Menon, Topfer, & Coloma, 1998). This is true for the older generation tricyclic antidepressants (TCAs; e.g. imipramine, desipramine, amitriptyline, clomipramine) as well as the more modern SSRIs (e.g., fluoxetine, fluvoxamine, paroxetine, sertraline, citalopram, venlafaxin) (Bet, Hugtenburg, Penninx, & Hoogendijk, 2013). Common side effects include, but are not limited to: insomnia, hypersomnia, nervousness, anxiety, agitation, tremor, restlessness, fatigue, somnolence, weight gain or weight loss, increased or decreased appetite, hypertension, sexual dysfunction, dry mouth, constipation, blurred vision, and sweating (Baldwin, 2003; Rosse, Fanous, Gaskins, & Deutsch, 2007). There is also growing evidence of treatment-emergent suicidal ideation, although the issue is still contentious (Laje et al., 2009; Perlis, Uher, Perroud, & Fava, 2012). Side effects vary across drugs (Bet et al., 2013; Rosse et al., 2007; Stahl, Grady, Moret, & Briley, 2005; Trindade et al., 1998), and some have more benign effects in specific domains. For instance, certain atypical antidepressants like mirtazapine (an α_2 adrenergic receptor antagonist) have been shown to have a superior sexual side effect profile (Serretti & Chiesa, 2009).

Side effects are common and persistent, especially during long-term antidepressants exposure (Bet et al., 2013). Adverse events during treatment increase significantly with antidepressant dosage (Bollini, Pampallona, Tibaldi, Kupelnick, & Munizza, 1999), and they are the number one reason for patients to discontinue treatment (Anderson & Tomenson, 1995). This is also the case for low-dose studies: a meta-analysis found that low-dose TCAs increased the likelihood of experiencing side effects by 50% compared to a placebo group, and dropout due to side effects was more than twice as likely in the treatment group (Furukawa, McGuire, & Barbui, 2002).

Interestingly, at least half of the common side effects listed above have one thing in common: they are the very symptoms that are supposed to *indicate* the presence of MDD. Clinical trials investigate the efficacy of drugs by assessing reductions of symptom sum-scores over time. The summands of the total scores, however – that are believed to indicate the presence of an underlying disorder – are partially exacerbated by treatment, seeing that antidepressants have adverse effects on MDD symptoms. Psychiatry thus finds itself in a highly curious situation. In December 2013, an email from the website alert.psychiatricnews.org contained an advertisement for Forfivo XL (bupropion), an atypical antidepressant (American Psychiatric Association, 2013b). Paradoxically, the advertisement encompassed the following safety information: "Antidepressants increased the risk compared to placebo of suicidal thinking and behavior (suicidality) […] in short-term studies of major depressive disorder (MDD) and other psychiatric disorders". On top of the fact that

antidepressants increase severity of symptoms that are then used to measure treatment efficacy across time, the instrument most commonly used in clinical trials is the HAM-D, which – compared to other depression screening scales such as the BDI – abounds in somatic symptoms that resemble the side effect profile caused by antidepressant treatment.

A detailed analysis of changes of individual symptoms in response to antidepressant treatment is likely to advance the field substantially by shedding light onto the ongoing debate about antidepressant efficacy from a new angle; currently, it is unclear if antidepressants are not efficacious, or if they improve some, yet worsen other symptoms.

4.5 Causal Associations Between Symptoms

Correlations between symptoms of depression are believed to be explained by the underlying common cause MDD, and should disappear once the latent is controlled for (see Chapter ›3.3). An alternative viewpoint has been put forward recently suggesting that symptoms have autonomous causal relevance (Borsboom & Cramer, 2013; Schmittmann et al., 2013) – that symptoms are directly and causally associated with each other beyond a latent variable that explains symptom covariance. To my knowledge, however, no literature overview has been conducted so far supporting this hypothesis. The present section gathers evidence for the causal power of two depression symptoms: insomnia and hopelessness.

Insomnia severely affects human functioning in various domains. It leads to psychomotor impairments (Dawson & Reid, 1997; Fairclough & Graham, 1999), cognitive impairments (Durmer & Dinges, 2005; Harrison & Horne, 2000), fatigue (Ferentinos et al., 2009), low mood (Murray, 1965; Reynolds et al., 1986; Samkoff & Jacques, 1991), and suicidal ideation or actual suicide (Fawcett et al., 1990; Sjöström, Waern, & Hetta, 2007). All of these symptoms either are or closely resemble DSM symptomatic criteria for depression (psychomotor problems; fatigue; diminished ability to think or concentrate, or indecisiveness; suicidal ideation). A meta-analysis of laboratory-based sleep loss studies exemplified the strength of these effects: sleep-deprived subjects performed 0.87 *SD* lower than the control group on psychomotor tasks, 1.55 *SD* lower on cognitive tasks, and reported mood 3.16 *SD* lower than the control group. When collapsing over all three measures, sleep-deprived subjects at the 50th percentile in their group performed equivalent to subjects at the 9th percentile in the control group (Pilcher & Huffcutt, 1996).

Hopelessness, on the other hand, describes negative expectancies about the future (Abramson, Metalsky, & Alloy, 1989)[6], and various studies have confirmed the predictive role of hopelessness for suicidal ideation and suicide (e.g., Fawcett et al., 1990, 1987). The effects are long-reaching: hopelessness predicted suicidal thoughts, attempts and actual suicide up to 13 years into the future in a large community sample (Kuo, Gallo, & Eaton, 2004) and was a predictor of suicide among psychiatric patients followed for up to 20 years (Beck, Brown, & Steer, 1989; Brown, Beck, Steer, & Grisham, 2000). Perhaps most interestingly, hopelessness was a better predictor of suicide than an inventory assessing all depressive symptoms in a prospective study of 1,958 outpatients with various psychiatric disorders (Beck, Brown, & Berchick, 1990): 94.2% of all individuals identified as high-risk group by the Beck Hopelessness Scale committed suicide, whereas 76.5% deemed high-risk by the BDI did. The same study also showed that the connection between hopelessness and suicide generalizes from individuals suffering from MDD to other psychiatric diagnoses, a notion that is supported by a recent study in which hopelessness predicted attempted suicide up to 5 years later in a sample of 414 psychotic patients, even after controlling for depression severity (Klonsky, Kotov, Bakst, Rabinowitz, & Bromet, 2012).

Causal relations between symptoms compromise the notion of locally independent and diagnostically interchangeable symptoms. However, many of the specific symptom-to-symptom pathways are controversial and merit further examination. While sleep deprivation, for instance, has been shown to have rapid mood-enhancing effects in depressed patients (Peterson & Benca, 2008), other reports suggest that insomnia causes low mood (Murray, 1965; Reynolds et al., 1986; Samkoff & Jacques, 1991). More studies are required to illuminate the complex effects of symptoms on other symptoms, taking into account situational, personal, and moderating variables.

Although incompatible with a reflective latent variable approach, the notion of symptoms that trigger, influence, or maintain other symptoms is widely accepted in clinical practice. A major goal in cognitive therapy is trying to break causal links between different MDD symptoms (Beck, Rush, Shaw, & Emery, 1979), and approaches like mindfulness-based cognitive therapy suggest that depressed mood leads to ruminative thinking, which in turn leads to other depressive symptoms (Ma & Teasdale, 2004; Nolen-Hoeksema, 2000).

---

[6] Although hopelessness is not part of the current diagnostic criteria, it is considered a depression symptom in this section for three reasons: first, it plays a major role in the cognitive triad originally described by Beck (1979); second, it performs more strongly than some DSM criterion symptoms in distinguishing depressed from healthy patients (McGlinchey et al., 2006); third, hopelessness is assessed in various MDD screening instruments as MDD symptom.

4.6 Symptoms as Predictors for Depression and Other Disorders

Covert heterogeneity documented in the last sections is contradictory of symptom interchangeability and underlines the clinical relevance of assessing and analyzing individual symptoms of depression. Yet another property that seems to be specific to individual symptoms – a property that is obfuscated by sum-scores – is the fact that certain symptoms are predictive of future episodes of depression as well as other disorders. Sleep problems serve as an example once more because they have received substantially more attention than other symptoms, and because they occur in a large proportion of depressed patients (Benca & Peterson, 2008; Peterson & Benca, 2008).

Ford and Kamerow (1989) were the first to note the importance of insomnia as a predictor of depression in a longitudinal study of 7,954 individuals; for those reporting insomnia at baseline, the probability to develop subsequent depression was drastically increased (odds ratio (OR) = 39.8), suggesting that early recognition and treatment of sleep problems may prevent depression onset. Since then, more than 40 studies have investigated the predictive value of sleep problems in depression (for a review, see Baglioni et al., 2011), some of which were longitudinal epidemiological reports (e.g., Breslau et al., 1996). Overall, non-depressed subjects with insomnia are at a twofold risk of developing depression in the future (Baglioni et al., 2011), and these findings roughly generalize to adolescents (OR = 2.3) (Roane & Taylor, 2008) and older adults (OR = 3.1) (Cho et al., 2008), and seem to be stable across different cultures (Okajima, Komada, Nomura, Nakashima, & Inoue, 2012). Additionally, depressed patients with persistent insomnia are more than twice as likely to remain depressed (Pigeon et al., 2008), and insomnia is a risk factor for relapse and recurrence (Perlis, Giles, Buysse, Tu, & Kupfer, 1997; Reynolds et al., 1997; Van Londen et al., 1998).

Further evidence for the clinical utility of individual depressive symptoms comes from studies that identified associations between specific depressive symptoms and other disorders. In a recently published prospective report of 54,279 men and women who were followed from 1995 through 2008, insomnia was predictive of heart failure (Laugsand, Strand, Platou, Vatten, & Janszky, 2013); compared to individuals without symptoms of insomnia, the ORs were 0.96, 1.35, and 4.53 for individuals reporting one, two, and three insomnia symptoms, suggesting that the evaluation of insomnia may be crucial for cardiovascular prevention. In another study, the MDD symptom loss of interest predicted Alzheimer's disease (Mossaheb et al., 2012): while 10.8% of those with a diagnosis at 60-month follow-up had shown loss of interest at baseline, only 2.2% in the non-demented group had (the symptom specificity in predicting Alzheimer's was 97.8, with a sensitivity of 10.4).

Individual depression symptoms may be important variables pertaining to prognosis and prevention of MDD and other disorders, and much more research is needed to understand the role of symptoms as predictors for the onset of mental and medical conditions. Unfortunately, it is unclear at present which specific symptom paths are responsible for increased risks to develop subsequent disorders. As discussed in Chapter ›4.5, insomnia is predictive of a variety of other depressive symptoms, and it is thus not surprising that insomnia also predicts depression (that is assessed via sum-scores) because insomnia itself is one of the summands of the sum-score. Disentangling which symptoms exactly influence other symptoms is among the most important epidemiological and statistical challenges of the next decade – a notion that likely generalizes from MDD research to other psychiatric diagnoses.

4.7 Summary

Numerous lines of evidence reveal profound heterogeneity within MDD that is commonly obscured by sum-scores, a notion that contrasts with the current disease model and the common cause hypothesis. However, comparably few prior studies aimed to uncover hidden symptom effects, and covert variability remains unexplored for three domains that are especially important.

First, while one previous cross-sectional study established differential associations of depressive symptoms with variables such as sex, history of depression, and personality traits (Lux & Kendler, 2010), the direction of causation has been unclear, and the question remains whether specific symptoms have different risk factors and etiologies. The study in Chapter ›5 addresses this problem by prospectively assessing a large sample of students before and after the onset of the chronic and intense stressor medical residency. The results reveal that increases of MDD symptoms are predicted by specific risk factors, suggesting that symptoms differ from each other in their etiologies. This offers crucial insights especially for the prevention of disease onset.

Second, while it is well-established that stress is a causal predictor of MDD, only a handful of prior studies have examined the associations between stress and individual depression symptoms (e.g., Keller & Nesse, 2005, 2006). These studies were cross-sectional, and it remains unclear how individual symptoms of depression change in response to severe life stress. The study in Chapter ›6 demonstrates marked differential increases of MDD symptoms following the onset of severe stress, again utilizing data from the prospective cohort study of medical residents. The results underline the importance of monitoring specific symptoms in high-risk populations.

Third, while depression is associated with substantial impairment of psychosocial functioning, only one previous report assessed concurrent effects of individual symptoms on functional

status (Tweed, 1993). Chapter ›7 extends the previous report in four important aspects: the study examines the differential impact of symptoms on impairment in a large and highly representative sample of depressed patients instead of a general population sample, uses the updated DSM-5 criterion symptoms instead of the DSM-III symptoms, investigates subsymptoms (e.g., hypersomnia and insomnia) instead of compound symptoms (e.g., sleep problems), and tests whether symptoms vary in their impacts across five different subdomains of impairment. The results demonstrate strong differential effects of symptoms on impairment; this means that two patients with similar symptom sum-scores may suffer from drastically different levels of functional impairment.

5 Study 1: Individual Depression Symptoms Have Different Risk Factors[7]

5.1 Abstract

Background. For diagnostic purposes, the nine symptoms that compose the DSM-5 criteria for MDD are assumed to be interchangeable indicators of one underlying disorder, implying that they should all have similar risk factors. The present study investigates this hypothesis, utilizing a population cohort that shifts from low to elevated depression levels.

Methods. We assessed the nine DSM-5 criterion symptoms for depression (using the Patient Health Questionnaire; PHQ-9) and seven depression risk factors (personal and family MDD history, sex, childhood stress, neuroticism, work hours, and stressful life events) in a longitudinal study of medical interns prior to and throughout internship (n = 1,289). We tested whether risk factors varied across symptoms, and whether a latent disease model could account for heterogeneity between symptoms.

Results. All MDD symptoms increased significantly during residency training. Four risk factors predicted increases in unique subsets of symptoms over time (depression history, childhood stress, sex, and stressful life events), while neuroticism and work hours predicted increases in all symptoms, albeit to varying magnitudes. MDD family history did not predict increases in any symptom. The pronounced heterogeneity of associations persisted after controlling for a latent depression factor.

Conclusions. The influence of risk factors varies substantially across DSM depression criterion symptoms. Since symptoms are etiologically heterogeneous, considering individual symptoms in addition to depression diagnosis might offer important insights obfuscated by symptom sum-scores.

---

[7] This chapter is published as: Fried, E. I., Nesse, R. M., Zivin, K., Guille, C., & Sen, S. (2013). Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors. *Psychological medicine*, Epub ahead of print. doi: 10.1017/S0033291713002900.

Accessible online: http://dx.doi.org/10.1017/S0033291713002900

5.2 Introduction

As reviewed in the introduction (see Chapter ›1.1), MDD is common, burdensome, recurrent, and expensive (Kessler et al., 2003, 2005; Lopez et al., 2006; Solomon et al., 2000). It is a highly heterogeneous disorder: 1,497 unique symptom profiles qualify for the same diagnosis, some of which do not share a single symptom (Ostergaard et al., 2011). Although variability in depression symptoms has been documented both across (Baumeister & Parker, 2012; Katschnig, Pakesch, & Egger-Zeidner, 1986) and within individuals (Coryell et al., 1994; Oquendo et al., 2004), the DSM and depression screening instruments such as the BDI and the HAM-D build sum-scores of depressive symptoms that are used to distinguish healthy from depressed individuals. This approach is based on the common cause hypothesis: since all symptoms are considered consequences of an underlying disease they can be considered interchangeable indicators.

Depression is also associated with a large number of risk factors. Identified risk factors for depression include demographic variables such as age and sex (Kendler, Myers, & Prescott, 2005; Piccinelli, 2000), personality traits such as neuroticism (Angst & Clayton, 1986; Kendler, Kuhn, & Prescott, 2004), early life adversity (Gilmer & McKinney, 2003; Gutman & Nemeroff, 2003), previous episodes of depression (Beekman et al., 1995; Colman et al., 2011), family history of depression (Nierenberg et al., 2007), and stressful life events (Mazure, 1998; Paykel, 2003).

If the common cause hypothesis is correct, then all depressive symptoms should have similar risk factors. This is because risk factors only influence the probability to develop the latent disorder, which in turn causes the symptoms. We propose an alternative hypothesis: risk factors differ for different depression symptoms. While this idea has been discussed in passing (Cramer et al., 2010; Hasler & Northoff, 2011; Gregor Hasler, Drevets, Manji, & Charney, 2004), it has not been directly tested before.

A recent comprehensive study investigated cross-sectional associations of all nine DSM criterion symptoms with 25 variables, including demographic information, personality traits, life events, history of depression, and lifetime comorbidities in a sample of 1,015 individuals (Lux & Kendler, 2010). The results revealed a complex association pattern. For instance, four symptoms were associated with years of education (sleep changes and fatigue with more education, psychomotor problems and suicidal ideation with less education), two symptoms were associated with family income (depressed mood with lower income, concentration problems with higher income), and two symptoms (depressed mood and psychomotor problems) had a significant positive correlation with current age. The authors concluded that the surprising degree of covert heterogeneity is

difficult to reconcile with the assumption of symptom equivalence. However, the study was cross-sectional, making a causal interpretation of association difficult. Furthermore, 225 uncorrected statistical tests were used to explore connections of variables, limiting the reliability of each individual finding.

A prospective investigation of a population that shifts from low to elevated depression levels allows a more direct examination of the influence of risk factors on specific symptoms. In the present study, we use medical residency as prospective stress model. Residency is a stage of U.S. graduate medical training, and usually includes a 1-year internship in which students work in a hospital. During this time interns face long work hours, sleep deprivation, loss of autonomy, as well as extreme emotional situations (Shanafelt & Habermann, 2002), making residency a well-established and intense stressor (Butterfield, 1988; Duffy, 2005). In a previous study of medical residents, depression levels increased from 3.9% at baseline to 25.7% during residency (Sen et al., 2010).

In summary, medical residents offers the rare opportunity to assess depressive symptoms and risk factors shortly before the onset of a severe and chronic stressor that drastically increases depression symptoms, allowing us to test whether increases in DSM depression criterion symptoms are predicted by different risk factors.

5.3 Methods

5.3.1 Sample

4,005 interns entering residency programs in the USA during the 2009-2011 academic years were invited to participate in the study; fifty-eight percent (2,455) accepted the invitation. The institutional review boards at participating hospitals approved the study. Participating subjects provided electronic informed consent, and were given $50 in gift certificates.

5.3.2 Assessment

All surveys were conducted through a secure online website designed to maintain confidentiality. Depressive symptoms were measured using the Patient Health Questionnaire (PHQ-9) (Spitzer, Kroenke, & Williams, 1999). The PHQ-9 is a self-report component of the PRIME-MD inventory designed to screen for the DSM-IV criterion symptoms of depression[8]. For each of the nine symptoms, subjects indicated whether, during the previous 2 weeks, the symptom had bothered them "not at all," "several days," "more than half the days," or "nearly every day." Each item yields a score of 0, 1, 2 or 3. The nine symptoms assessed by the PHQ-9 are: little interest or pleasure in

[8] DSM-IV and DSM-5 criterion symptoms for MDD are identical.

doing things *(interest)*, feeling depressed or hopeless *(depressed)*, sleep problems *(sleep)*, feeling tired *(fatigue)*, appetite problems *(appetite)*, feeling bad about yourself / that you are a failure *(self-blame)*, trouble concentrating on things *(concentration)*, moving or speaking slowly / being fidgety or restless *(psychomotor)*, and suicidal ideation *(suicide)*.

Subjects completed a baseline survey 1-2 months prior to commencing internship that assessed general demographic factors (age, sex), personal factors (baseline PHQ-9 depressive symptoms, self-reported history of depression, self-reported family history of depression, and childhood stress (Risky Families Questionnaire; Taylor et al., 2006)), and neuroticism (NEO-Five Factor Inventory; Costa & McCrae, 1997).

Participants were contacted via email at months 3, 6, 9 and 12 of their internship year and asked to complete the PHQ-9 again. They were also queried regarding work hours over the past week and the occurrence of a series of stressful non-internship life events (marriage, childbirth, serious illness, death or serious illness in close family or friend, financial problems, end of a serious relationship, or becoming a victim of crime or domestic violence) during the past 3 months. Because the number of life events reported in each category was too low to use in subsequent analyses, we computed a sum-score of all events for each measurement point per subject and subsequently used it as predictor.

5.3.3 Statistical Analysis

We compared symptom severity at baseline with average symptom severity during the four measurements across the residency. This approach has been used in previous publications based on this dataset (e.g., Sen et al., 2010, 2013), and has the advantage of increased reliability of symptom assessment within-internship through repeated measurement. When averaging the within-internship symptom scores, 1,166 of the 2,455 subjects (47.5%) were dropped via listwise deletion because they had missing data on two or more timepoints, leaving 1,289 subjects in the analytic sample. Average scores for the two risk factors that were assessed after baseline (work hours and number of stressful life events) were constructed analogous to symptoms.

Overall, three analyses were conducted. First, we used a paired *t*-test to compare baseline levels of the PHQ-9 to the average PHQ-9 level during residency[9].

Second, we investigated whether risk factors differentially predicted symptom change over time. We estimated a structural equation model (SEM) encompassing nine linear regressions, one regression per symptom. In each regression, severity of a given symptom during residency was

[9] Detailed information on individual symptom increases can be found in Study 2 (see Chapter ›6).

predicted by five baseline risk factors (sex, history of depression, family history of depression, neuroticism, and childhood stress) and two within-internship risk factors (work hours and number of stressful events), controlling for baseline severity of the symptom. To avoid overfitting, we used identical predictors in all nine regressions, and did not drop insignificant predictors to reach the best possible fit for each symptom; that is, if a particular risk factor did not predict increases on a certain symptom, the predictor was retained nonetheless in this specific regression. However, we excluded predictors from the analysis in case they were not related to *any* of the nine symptoms. We then compared two models: in the *homogeneity* model, regression weights were constrained to be equal for each risk factor across all symptoms. Specifically, a regression coefficient was estimated for each risk factor separately, but this coefficient was constrained to have the same effect on each symptom. The homogeneity model represents the hypothesis that depression is the common cause for all symptoms, with each risk factor affecting all symptoms equally. In the *heterogeneity* model, all regression coefficients were freely estimated. The models were compared using a χ^2 test, and model fit was assessed using the root mean square error of approximation (RMSEA; RMSEA ≤ 0.06 indicating good fit) as well as the comparative fix index (CFI; CFI ≥ 0.95 indicating good fit) (Hu & Bentler, 1999). If the constrained homogeneity model would fit worse than the heterogeneity model, this would indicate that risk factors vary in their effects across symptoms. Both models were estimated using the Maximum Likelihood estimator (ML), and results of the model with better fit were visualized using the R-package QGRAPH (Epskamp et al., 2012).

While our second analysis provides information about potential heterogeneity on the basis of observable symptoms, it does not rule out the possibility that heterogeneity disappears in a latent disease model. Thus, we investigated whether controlling for a latent depression variable eliminates heterogeneity, using a longitudinal multiple-indicator multiple-cause (MIMIC) model. MIMIC models (Jöreskog & Goldberger, 1975) encompass one or more latent variables that have both a number of observable indicators (in our case symptoms) and variables that influence the latent (in our case risk factors). Following the approach described by Jones (2006), we compared two versions of a MIMIC model. The first model (model I) allows for risk factors to directly influence symptoms, while the second model (model II) represents the hypothesis that risk factors only affect the latent depression variable which, in turn, influences individual depressive symptoms (Figure 5-1). Evidence that these direct paths improve model fit significantly would indicate that symptom level effects of risk factors exist and are not mediated by a latent variable. Model I was constructed in an iterative process. First, all risk factors were allowed to have direct effects on all symptoms, except for the last PHQ-9 symptom suicidal ideation (for the purpose of model identi-

fication). Second, non-significant direct paths were removed and significant ones added for suicidal ideation until only significant paths from risk factors to symptoms remained. Model comparison and fit of the two MIMIC models were assessed analogous to analysis two. A complication to the MIMIC model is the longitudinal nature of our data. Similar to our second analysis, we thus predicted the second measurement point by risk factors, controlling for the first. While work hours and number of stressful life events that were assessed during residency naturally only affected the within-internship latent, baseline risk factors (sex, history of depression, neuroticism, and childhood stress) were allowed to additionally affect the baseline latent variable. In both models, the residual of each symptom at baseline was allowed to be correlated with the residual of the same symptom during residency. Both MIMIC models were estimated using the ML estimator[10].

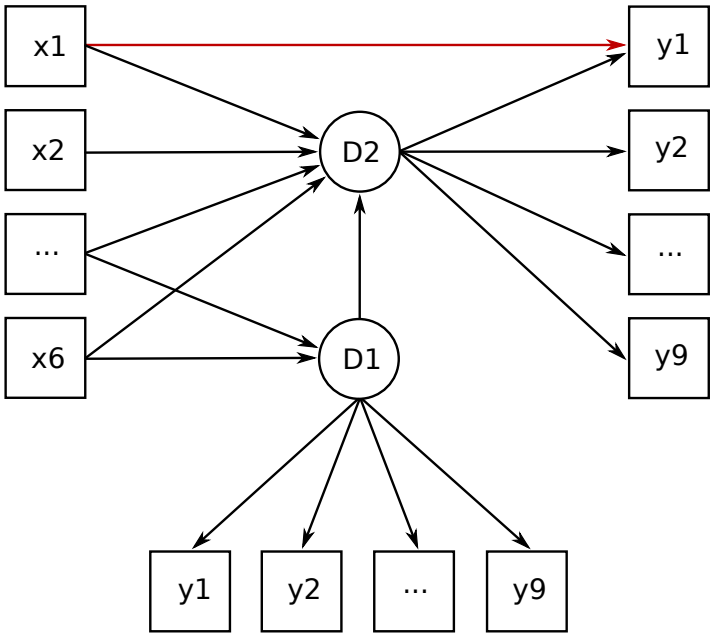

Figure 5-1. *Visualization of longitudinal MIMIC models.*
D1, latent depression factor at baseline; D2, latent depression factor during residency; y1-y9, depressive symptoms; x1-x2, within-internship risk factors; x3-x6, baseline risk factors; the red arrow is an example of a significant direct effect of a risk factor on a symptom (in this case, x1 on y1) that is not mediated by D2.

Analysis one was performed using R v2.13.0 (R Development Core Team, 2008), analyses two and three were conducted in MPLUS v7.0 (Muthén & Muthén, 2012).

[10] To confirm the accuracy of the ML estimator we repeated the analysis using the Maximum Likelihood estimator with robust standard errors (MLR), bootstrapping the standard errors (5,000 iterations). As the results were essentially unchanged, we retained the original, conceptually simpler models.

5.4 Results

1,289 interns were included in the final analyses. The mean age of study participants was 27.6 (*SD* = 3.02) and 47.9% were male (see Table 5-1 for detailed characteristics). Participants that had to be dropped due to missing values did not differ significantly from the retained participants in sex, neuroticism, childhood stress, personal and family history of depression, work hours, and number of stressful life events (all $p > 0.05$).

5.4.1 Effects of residency on depression

The PHQ-9 score increased from 2.37 points (*SD* = 2.96) at baseline to 5.73 points (*SD* = 3.94) within-residency ($p < 0.001$).

Variable	Number (%)	Variable	Number (%)
Sex		Specialty	
Male	617 (47.9)	Internal Medicine	481 (37.3)
Female	672 (52.1)	Other	158 (12.3)
Age, y		Pediatrics	150 (11.6)
≤ 25	221 (17.1)	General Surgery	113 (8.8)
26-30	867 (67.3)	Psychiatry	93 (7.2)
31-35	111 (8.6)	Emergency Medicine	81 (6.3)
> 35	28 (2.2)	Family Medicine	62 (4.8)
History of depression		Obstetrics/gynecology	47 (3.6)
Yes	568 (44.1)	Internal medicine/pediatrics	35 (2.7)
No	721 (55.9)	Neurology	32 (2.5)
MDD family history		Transitional	27 (2.1)
Yes	682 (52.9)	Missing	10 (0.8)
No	607 (47.1)		

Table 5-1. *Demographic characteristics of study participants.*

5.4.2 Impact of risk factors on symptoms

Family history of depression was the only risk factor that was not related to changes of *any* of the nine symptoms and was excluded from subsequent analyses. The homogeneity model in which each risk factor was constrained to have equal impact on all symptoms fit the data significantly worse than the heterogeneity model in which the effects of risk factors on symptoms were freely estimated ($p < 0.001$). The heterogeneity model fit the data well, and the highly significant χ^2 difference test indicated strong differential impact of risk factors on symptoms (see Table 5-2).

	χ^2	*df*	RMSEA	CFI	χ^2_{diff}	df_{diff}	*p*
Test for heterogeneity							
Model I[a]	197.3	72	0.04	0.98			
Model II[b]	617.4	120	0.06	0.93	420.1	48	< 0.001
MIMIC models							
Model I[c]	868.5	199	0.05	0.94			
Model II[d]	1041.5	218	0.05	0.93	173.0	19	< 0.001

Table 5-2. *Goodness-of-fit statistics and χ^2 difference tests for the two model comparisons.*
df, degrees of freedom; RMSEA, root mean square error of approximation; CFI, comparative fit index; χ^2_{diff}, χ^2 statistic of the χ^2 difference test; df_{diff}, degrees of freedom of the χ^2 difference test; *p*, *p*-value of the χ^2 difference test; a, heterogeneity model; b, homogeneity model; c, 19 significant direct paths from risk factors to symptoms; d, no direct paths from risk factors to symptoms.

Full results of the heterogeneity model are summarized in Table 5-3, and visualized in Figure 5-2. Of the various findings, three are particularly worthy of note. First, four risk factors (prior history of depression, higher number of stressful life events, childhood stress, and sex) predicted worsening of unique subsets of three to seven of the symptoms. Second, two predictors (neuroticism and work hours) had significant impact on all nine PHQ-9 symptoms over time. Third, female residents showed increases in sleep problems, appetite problems, and fatigue during residency, while male residents reported increased suicidal ideation.

	Female sex	Neuroticism	Childhood stress	History of depression	Stressful life events	Work hours	R^2
Interest	-0.01	0.29 ***	0.05	0.13 ***	0.05	0.15 ***	0.20
Depressed	0.01	0.32 ***	0.04	0.15 ***	0.03	0.09 ***	0.23
Sleep	0.09 ***	0.20 ***	0.02	0.09 **	0.04	0.07 **	0.13
Fatigue	0.10 ***	0.20 ***	0.05	0.11 ***	0.05 *	0.24 ***	0.20
Appetite	0.09 ***	0.24 ***	0.04	0.08 **	0.05 *	0.15 ***	0.20
Self-blame	0.04	0.36 ***	0.03	0.11 ***	0.02	0.10 ***	0.26
Concentration	0.05 *	0.19 ***	0.08 **	0.12 ***	0.08 **	0.14 ***	0.17
Psychomotor	-0.04	0.22 ***	0.08 **	0.03	0.06 *	0.14 ***	0.11
Suicide	-0.08 **	0.22 ***	0.10 ***	0.04	0.08 **	0.05 *	0.14

Table 5-3. *PHQ-9 symptoms predicted by risk factors (heterogeneity model).*
Standardized regression coefficients of risk factors on symptoms. R^2, adjusted *R*-squared values. Family history of depression is not displayed because none of the effects was significant. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

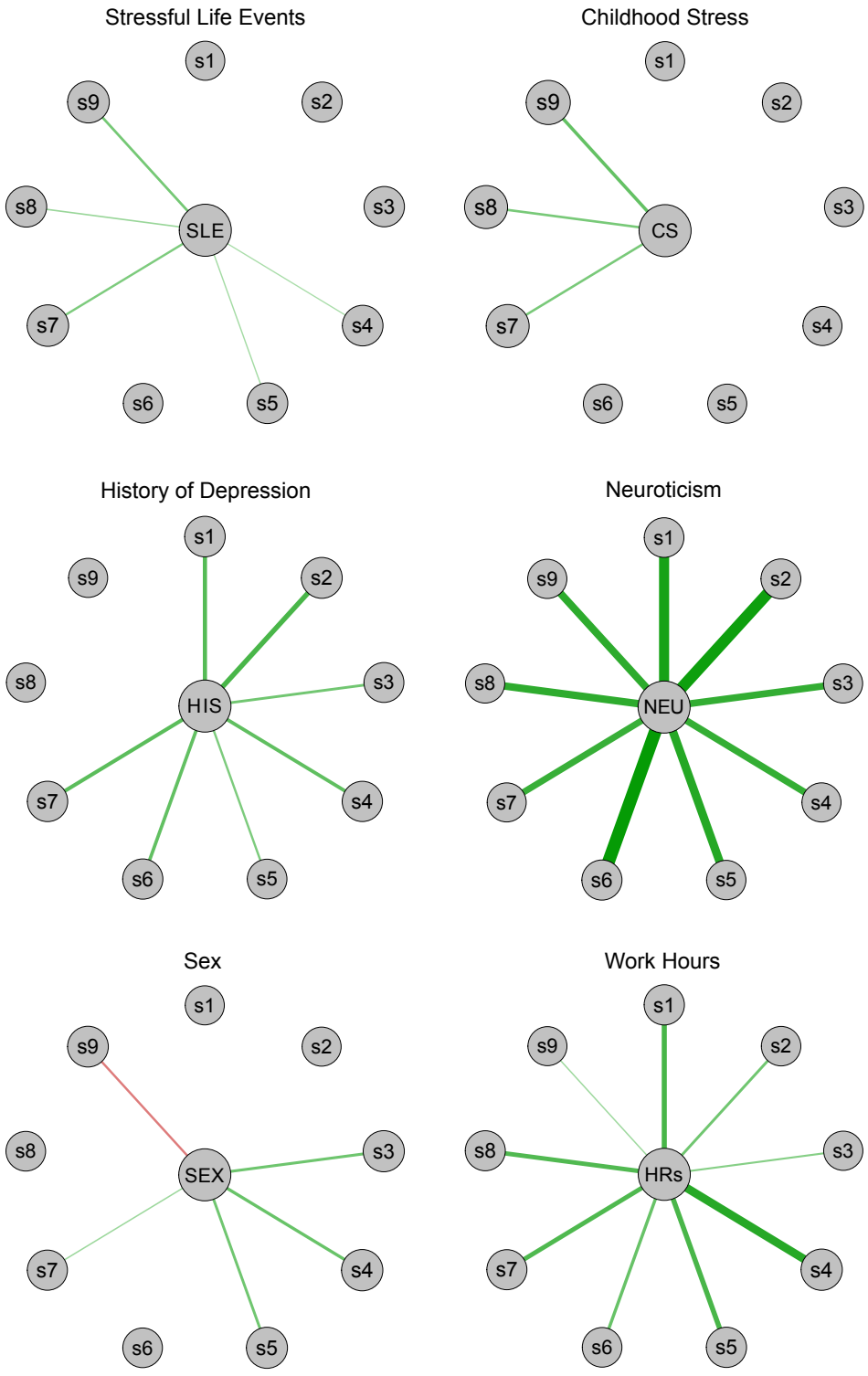

Figure 5-2. *Regression coefficients of risk factors on changes of depression symptoms (heterogeneity model).* Thickness of lines indicates strength of standardized regression weights. Green lines represent positive regression weights, red lines negative ones; sex was coded 0=male, 1=female. s1, interest; s2, depressed; s3, sleep; s4, fatigue; s5, appetite; s6, self-blame; s7, concentration; s8, psychomotor; s9, suicide.

To assess whether the two uniform predictors neuroticism and work hours varied in their predictive strength across symptoms, we compared confidence intervals (CI) of the regression coefficients across symptoms. Of 36 symptom pairs, 14 CI (38.9%) did not overlap for neuroticism, and 14 CI (38.9%) did not overlap for work hours (Table 5-4). This implies considerable variability in the predictive strength of neuroticism and work hours across symptoms.

		Neuroticism	Work hours
Interest	*B*	0.16	0.09
	CI	0.13-0.20	0.06-0.11
Depressed	*B*	0.18	0.05
	CI	0.15-0.21	0.02-0.08
Sleep	*B*	0.14	0.05
	CI	0.10-0.18	0.01-0.08
Fatigue	*B*	0.13	0.16
	CI	0.10-0.17	0.13-0.19
Appetite	*B*	0.18	0.11
	CI	0.14-0.22	0.08-0.15
Self-blame	*B*	0.22	0.06
	CI	0.19-0.26	0.03-0.09
Concentration	*B*	0.12	0.08
	CI	0.08-0.15	0.05-0.11
Psychomotor	*B*	0.09	0.06
	CI	0.07-0.11	0.04-0.08
Suicide	*B*	0.06	0.01
	CI	0.04-0.07	>0.00-0.03

Table 5-4. *Regression coefficients and confidence intervals for neuroticism and work hours.*
B, unstandardized regression coefficients; *CI*, 95% confidence intervals of the regression coefficients.

5.4.3 Latent disease model

In this analysis, we compared the fit of two longitudinal MIMIC models. Since preliminary analyses revealed that factor loadings were not invariant across time, we allowed factor loadings to vary across measurement points in both MIMIC models. Due to high modification indices, residuals of the following symptoms were allowed to be correlated in both models: symptoms sleep and fatigue at baseline, and symptoms depressed and appetite, depressed and self-blame, fatigue and appetite, as well as concentration and psychomotor within-internship. As result of the iterative fitting process described in the Methods section, model I retained 19 significant direct effects of risk factors on symptoms after controlling for the latent depression factor.

Results of fitting the two models to the data are presented in Table 5-2. The χ^2 difference test indicated that the 19 direct paths from risk factors to symptoms in model I improved the fit significantly ($p < 0.001$). This means that heterogeneity of risk factors and symptoms cannot be explained by a latent depression variable.

5.5 Discussion

In this study, we tested the hypothesis that the nine DSM-5 depression symptoms have similar risk factors. The evidence from a longitudinal population cohort of 1,289 medical interns is inconsistent with this hypothesis. Instead, the results support the alternative hypothesis that risk factors have differential impact on depression symptoms. Four risk factors predicted increases of unique subsets of symptoms. Neuroticism and work hours did predict increases of all symptoms, albeit to varying magnitudes. We also tested whether the pronounced heterogeneity could be explained by a latent depression factor; this was not the case.

These results support the notion that depressive symptoms have different characteristics and properties. Studies of depression typically use sum-scores to describe severity, a strategy that may obfuscate crucial information about the nature of depression symptoms and causes. For instance, in this report, women were more likely to report worsening of sleep and appetite problems as well as fatigue during internship, whereas men reported increased suicidal ideation – information that would not have been available from assessing depression sum-scores alone. Furthermore, the use of sum-scores groups individuals with dramatic inter-individual symptom differences together in the same diagnostic group.

Overall, our results are consistent with the growing chorus of voices suggesting that covert heterogeneity of DSM diagnostic criteria in research may be inhibiting progress in elucidating the biological roots of mental illness (Insel, 2013; Kendler et al., 2013; Lux & Kendler, 2010). Research informed about differences between individual symptoms will facilitate the search for etiologically and symptomatically homogenous groups, and thus serve our main goal of increasing treatment efficacy.

5.6 Limitations

These findings need to be interpreted in the light of four limitations. First, medical residents are relatively homogenous in terms of age, education, and SES, so our results may not generalize to other populations.

Second, the PHQ-9 does not contain detailed information about bi-directional symptoms. It is thus unclear whether, for instance, the increasingly severe sleep problems in female residents manifested themselves as insomnia or hypersomnia, or whether the increases of psychomotor problems reported by interns that experienced a higher number of stressful life events were due to psychomotor agitation or psychomotor retardation.

Third, in preliminary analyses conducted prior to MIMIC modelling, a one-factor solution emerged for the first measurement point, while two factors fit the second measurement point better. Seeing that the MIMIC models were not interpretable from a substantive point of view with different numbers of factors across time, we estimated one factor per measurement point, but allowed factor loadings to be freely estimated. This approach penalizes the fit of both MIMIC models equally and should not favor one over the other.

Fourth, because multivariate models that contained nine dependent variables, six risk factors, five measurement points, and controlled for a latent depression factor were not interpretable, our analyses are limited to two timepoints. As a result, if the relationship between specific variables changed during internship, these changes may have been masked. However, correlations of symptoms with time-varying risk factors at each individual timepoint were fairly stable across the four post stress onset measurements, and comparable to the averaged timepoint used in our analyses. Our approach therefore does not seem to substantially distort the relationships of these variables.

5.7 Acknowledgments

We would like to thank all interns who participated in the study for their kind cooperation.

6 Study 2: The Differential Influence of Life Stress on Individual Symptoms of Depression[11]

[11] This chapter was submitted to *The British Journal of Psychiatry:* Fried, E. I., Nesse, R. M., Guille, C., & Sen, S.: "The differential influence of life stress on individual symptoms of depression". Due to copyright reasons the chapter is not available in the online version of this dissertation, please contact the corresponding author at eiko.fried@gmail.com for a copy of the manuscript.

7 Study 3: The Impact of Individual Depressive Symptoms on Impairment of Functioning[17]

7.1 Abstract

Background. Previous studies have established that scores on MDD scales are correlated with measures of impairment of psychosocial functioning. It remains unclear, however, whether individual depressive symptoms vary in their effect on impairment, and if so, what the magnitude of these differences might be.

Methods. We analyzed data from 3,703 depressed outpatients in the first treatment stage of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study. Participants reported on the severity of 14 depressive symptoms, and stated to what degree their depression impaired psychosocial functioning (in general, and in the five domains work, home management, social activities, private activities, and close relationships). We tested whether symptoms differed in their associations with impairment, estimated unique shared variances of each symptom with impairment to assess the degree of difference, and examined whether symptoms had variable impacts across impairment domains.

Results. Our results show that symptoms varied substantially in their associations with impairment ($p < 0.001$), and contributed to the total explained variance in a range from 0.7% (hypersomnia) to 20.9% (sad mood). Furthermore, symptoms had variable impacts across the five impairment domains ($p < 0.001$). Overall, sad mood and concentration problems had the highest unique associations with impairment and were among the most debilitating symptoms in all five domains.

Conclusions. Our findings are in line with a growing chorus of voices suggesting that symptom sum-scores obfuscate relevant differences between depressed patients and highlight the potential utility of considering individual symptoms in addition to depression diagnosis.

[17] This chapter is accepted as: Fried, E. I., & Nesse, R. M. (2014). The Impact of Individual Depressive Symptoms on Impairment of Psychosocial Functioning. *PLOS ONE*.

7.2 Introduction

About 60% of individuals who meet criteria for MDD as defined by the DSM-5 report severe or very severe impairment of functioning (Kessler et al., 2003). Impairment associated with depression is long-lasting (Hays et al., 1995) and equal or greater than impairment caused by other common, chronic medical conditions such as diabetes, hypertension, heart attack, and congestive heart failure (Mathers & Loncar, 2006; Murray & Lopez, 1996). Depression impairs functioning in different domains such as home life, workplace, and family (Hirschfeld et al., 2002; Judd et al., 1996), and often severely compromises the capacity for self-care and independent living.

A recent review found scores on various screening instruments for depression to be moderately correlated with measures of impairment (McKnight & Kashdan, 2009). It has been unclear, however, whether certain symptoms are more impairing than others, and if so, what the magnitude of these differences might be. This question is highly relevant because of the large symptomatic differences in MDD, both across individuals diagnosed with depression (Katschnig et al., 1986; Lichtenberg & Belmaker, 2010; Ostergaard et al., 2011) and within individuals across time (Coryell et al., 1994; Oquendo et al., 2004).

We are aware of only a single previous study that explored concurrent effects of individual depressive symptoms on impairment of psychosocial functioning (Tweed, 1993). In this analysis of a general population sample, six DSM-III symptoms were significantly associated with impairment (depressed mood, dysthymia, cognitive difficulties, suicidal ideation, fatigue, and sexual disinterest). The present study extends the previous report (Tweed, 1993) in five important aspects. First, we examine the differential impact of symptoms on impairment in a large and highly representative sample of 3,703 depressed patients, instead of a general population sample. Second, in addition to estimating which symptoms are significantly associated with impairment, we employ a statistical approach that provides a detailed estimation of the degree of these associations. Third, we use the updated DSM-5 criterion symptoms instead of the DSM-III criteria. Fourth, we investigate subsymptoms (e.g., psychomotor agitation and psychomotor retardation) instead of compound symptoms (e.g., psychomotor problems). Lastly, we test whether symptoms vary in their impacts across five different domains of impairment.

7.3 Methods

7.3.1 Study description

Data from the first treatment stage (level 1) of the NIH-supported "Sequenced Treatment Alternatives to Relieve Depression" (STAR*D) study (Fava et al., 2003; Rush et al., 2004) were analyzed for this report. The authors obtained NIMH Data Use Certificates to use the STAR*D datasets (version 3). STAR*D was a multisite randomized clinical trial conducted in the USA to investigate which of several treatment options would be most effective for nonpsychotic MDD outpatients; 4,041 patients were enrolled into the first treatment stage, in which all participants received citalopram (SSRI). Outcome data were obtained via telephone interviews that were conducted either by interviewers, or by an interactive voice response system (IVR). STAR*D was approved and monitored by the institutional review boards at each of the 14 participating institutions, a national coordinating center, a data coordinating center, and the data safety and monitoring board at the NIMH. All participants provided written informed consent at study entry. Further more specific information about design, methods, exclusion criteria, and the rationale of STAR*D can be obtained elsewhere (Fava et al., 2003; Rush et al., 2004).

7.3.2 Participants

STAR*D used relatively inclusive selection criteria in order to obtain a highly representative sample of patients seeking treatment for MDD. Participants had to be between 18 and 75 years of age, fulfill DSM-IV criteria for single or recurrent nonpsychotic MDD, and have at least moderately severe depression corresponding to a score of at least 14 on the 17-item HAM-D. Participants with a history of bipolar disorder, schizophrenia, schizoaffective disorder, or psychosis were excluded, as were patients with current anorexia, bulimia, or primary obsessive-compulsive disorder. Further exclusion criteria were a history of intolerability to antidepressant medication, lack of response to an adequate trial of SSRI in the current episode of MDD, or failure to respond to 16 or more sessions of cognitive therapy in the current episode of MDD. Our analyses are limited to the 3,703 individuals that were assessed within the first week of level 1 via IVR.

7.3.3 Outcomes measures

7.3.3.1 Depressive symptoms

STAR*D used the Quick Inventory of Depressive Symptoms (QIDS-16; Rush et al., 2003) to assess depressive symptoms. The QIDS-16 has good psychometric properties (Rush et al., 2003), and the results of the IVR version are comparable to the results produced by the self-rated and the

clinician-rated QIDS-16 (Rush et al., 2006). The QIDS-16 assesses the nine DSM symptom domains with 16 questions. Each domain yields a score between 0 and 3, 0 indicating no problems, 3 indicating severe problems. While six symptoms are measured with single questions, the three compound symptoms *(sleep problems*, *psychomotor problems*, *appetite/weight problems)* are assessed with multiple questions. The QIDS-16 constructs these compound symptoms by using the highest symptom score in each symptom group, resulting in one score on each of the nine DSM criterion symptoms. Since we were interested in individual symptoms, we used all available items instead of symptom domains. Detailed information for the domain *appetite and weight problems* was not available, since either *appetite decrease* or *appetite increase*, and either *weight decrease* or *weight increase* was scored. Overall, this resulted in twelve individual symptoms plus the two compound symptoms *appetite problems* and *weight problems* (Table 7-1).

QIDS-16 symptoms	Shortcode
Sleep onset insomnia	Early insomnia
Mid-nocturnal insomnia	Middle insomnia
Early morning insomnia	Late insomnia
Hypersomnia	Hypersomnia
Sad Mood	Sad mood
Appetite increase	Appetite
Appetite decrease	Appetite
Weight increase	Weight
Weight decrease	Weight
Problems concentrating / making decisions	Cognition
Feeling worthless / self-blame	Self-blame
Suicidal ideation	Suicidal ideation
Loss of interest	Interest loss
Energy loss / fatigability	Fatigue
Psychomotor slowing	Slowed
Psychomotor agitation	Agitated

Table 7-1. *QIDS-16 depressive symptoms.*

We chose to analyze the QIDS-16 instead of other measures of depressive symptoms also assessed in STAR*D for several reasons. First, the QIDS-16 is the only depression scale that was assessed with the same modality as measures of impairment (IVR). Second, the QIDS-16 contains all DSM-5 criterion symptoms for depression, including detailed information on two of the three bi-directional compound symptoms. Third, when investigating effects of symptoms on debilitation it is paramount not to artificially inflate shared variance between specific symptoms and impairment. This is likely to be the case when one uses depressive symptom scales that assess impairment, or impairment scales that assess depressive symptoms; for instance, the HAM-D, also assessed in

STAR*D, includes an item measuring impairment (item 10). Fourth, symptoms must be scored on the same scale to allow for proper comparisons between symptoms. All QIDS-16 items are scored on a scale ranging from 0-3, while the HAM-D, for instance, uses ranges from 0-2 to 0-4 for symptoms.

7.3.3.2 Impairment

The Work and Social Adjustment Scale (WSAS; Marks, 1986) was used to measure impairment of functioning. The WSAS is a simple, reliable, and valid self-report instrument that uses Likert-scale ratings of 5 items to assess impairment in the domains of work, home management, social activities, private activities, and close relationships. Each question is rated on a 0-8 Likert scale, with 0 indicating no impairment and 8 indicating very severe impairment. WSAS scores below 10 are associated with subclinical populations; scores of 10-20 are associated with significant functional impairment, while scores above 20 suggest at least moderately severe functional impairment (total range 0-40). The WSAS has been used predominantly in samples with mood and anxiety disorders, and has been shown to have good internal consistency (0.70 to 0.94) and retest-reliability (0.73), and high concurrent validity of IVR administrations with clinician interviews (0.81 and 0.86) (Mundt, Marks, Shear, & Greist, 2002). In STAR*D, the WSAS specifically queried participants how much *their depression* impaired work and social activities. For instance, work impairment was measured via the following item: "Because of my depression, my ability to work is impaired. 0 means not at all impaired and 8 means very severely impaired to the point I can't work."

The WSAS was chosen over other measures of impairment used in STAR*D for several reasons. First, the WSAS assesses both overall impairment as well as impairment in several subdomains. Second, the WSAS version used in STAR*D specifically instructed participants to report how depression interfered with their functioning, in contrast to the other instruments that either assessed effects of health problems on impairment (Work Productivity and Activity Impairment Questionnaire (WPAI); Reilly, Zbrozek, & Dukes, 1993), or did not query individuals about the effects of depression on reduced functioning (Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q), Endicott, Nee, Harrison, & Blumenthal, 1993; Short-Form Health Survey (SF-12), Ware & Sherbourne, 1992). Third, as mentioned above, it is crucial to minimize shared variance between depressive symptoms and impairment scales that could potentially bias the results. Both the Q-LES-Q (item 2) and the SF-12 (item 11) assessed mood as one of the items of impairment, potentially inflating the importance of affective depression symptoms on impairment, especially depressed mood.

7.3.4 Statistical analysis

Three analyses were performed. First, we used the 14 QIDS-16 depression symptoms to predict overall impairment as measured by the WSAS sum-score, controlling for age and sex. We then compared two linear regression models: in model I (heterogeneity model), regression weights for symptoms were free to vary, whereas model II (homogeneity model) constrained regression weights to be equal. While model I allows for differential impairment-symptoms associations, model II represents the hypothesis that symptoms have equal associations with impairment. A χ^2-test was used to compare the two models. Because depressive symptoms are generally correlated with each other, we performed multicollinearity diagnostics for both regression analyses. The variance inflation factor (VIF) did not exceed the value of five for any symptom, indicating no multicollinearity problems (Heiberger & Holland, 2004).

Second, we aimed to allocate unique R^2 shares to each regressor to examine how much unique variance each individual symptom shared with impairment. We used the LMG metric via the R-package RELAIMPO (Grömping, 2006) to estimate the relative importance (RI; see Grömping, 2007; Johnson & Lebreton, 2004; Kruskal & Majors, 1989) of each symptom. LMG estimates the importance of each regressor by splitting the total R^2 into one non-negative R^2 share per regressor, all of which sum to the total explained R^2. This is done by calculating the contribution of each predictor at all possible points of entry into the model, and taking the average of those contributions. In other words, an estimate of RI for each variable is obtained by estimating as many regressions as there are possible orders of regressors (in the present case, 8.7×10^{10} regressions[18]), and then averaging individual R^2 values over all models. RI estimates are then adjusted to sum to 100% for easier interpretation. Confidence interval (CI) estimates of the RI coefficients, as well as *p*-values indicating whether regressors differed significantly from each other in their RI contributions (in an exploratory sense), were obtained using the bootstrapping capabilities of the RELAIMPO package. It is important to note that predictors with a non-significant regression coefficient can nonetheless contribute to the total explained variance, that is, have a non-zero LMG contribution. This is the case when regressors are correlated with each other and thus can indirectly influence the outcome via other regressors (Grömping, 2007). Therefore, all symptoms, even those with non-significant coefficients, were included in subsequent RI calculations.

Third, we tested whether individual symptoms differed in their associations across the five WSAS impairment domains work, home management, social activities, private activities and close

[18] 14 choose 14 (ordered with no repeats): $14! / (14-14)!) = 8.7 \times 10^{10}$.

relationships. We estimated two structural equation models (SEM), using the ML estimator. Both models contained five linear regressions, one for each domain of impairment. In each of these five regressions, we used the 14 depressive symptoms as predictors, controlling for age and sex. While the first SEM allowed free estimation of all regression coefficients (model I), the second constrained each symptom to have equal effects across the five impairment domains (model II). This second model represents the hypothesis that a given symptom has similar impacts on all five domains. We compared the models using a χ^2-test. The results of the better fitting model were visualized using the R-package QGRAPH (Epskamp et al., 2012).

Analyses one and three were performed in MPLUS v7.0 (Muthén & Muthén, 2012), and analysis two was estimated in R v2.13.0 (R Development Core Team, 2008).

7.4 Results

Of the 3,703 outpatients in the study, 2,234 (60.3%) were female, and the mean age was 41.2 years ($SD = 13.2$). See Table 7-2 for detailed demographic information.

Variable	Number (%)	Variable	Number (%)
Age, y		Marital Status	
≤ 20	86 (2.3)	Never married	1091 (29.5)
21-30	842 (22.7)	Cohabitating with partner	310 (8.4)
31-40	835 (22.5)	Married	1238 (33.4)
41-50	915 (24.7)	Separated	245 (6.6)
51-60	711 (19.2)	Divorced	698 (18.8)
> 60	314 (8.5)	Widowed	117 (3.2)
Race		Missing	4 (0.1)
White	2926 (79.0)	Employment status	
Black/African American	685 (18.5)	Unemployed	1379 (37.3)
Other	92 (2.5)	Employed	2101 (56.8)
Ethnicity: Hispanic	452 (12.2)	Retired	218 (5.9)
		Missing	5 (0.1)

Table 7-2. *Demographic characteristics of study participants.*

The average impairment score was 23.52 ($SD = 9.29$), corresponding to moderately severe levels of impairment; 307 (8.3%) individuals did not show impaired functioning, 875 (23.6%) exhibited significant functional impairment, while 2,521 (68.1%) reported severe functional impairment.

7.4.1 Homogeneity versus heterogeneity of associations

The heterogeneity model (allowing variable contributions of symptoms to impairment) fit the data significantly better than the homogeneity model (in which symptoms were constrained to have the

same contributions to impairment) ($\chi^2 = 394.5$, $df = 13$, $p < 0.001$). In the heterogeneity model, 11 of the 14 depression symptoms significantly predicted impairment, explaining 40.8% of the variance (F (16, 3686) = 159.1, $p < 0.001$) (Table 7-3). The heterogeneity model was thus used for subsequent RI estimations.

Predictors	*b*	*s.e.*	*t*	
Early insomnia	0.50	0.11	4.53	***
Middle insomnia	0.01	0.15	0.08	
Late insomnia	0.26	0.11	2.32	*
Hypersomnia	0.54	0.15	3.64	***
Sad mood	2.27	0.18	12.79	***
Appetite	0.25	0.12	2.14	*
Weight	0.13	0.11	1.17	
Cognition	1.61	0.14	11.21	***
Self-blame	0.68	0.10	6.61	***
Suicidal ideation	0.84	0.15	5.50	***
Interest loss	1.24	0.12	10.40	***
Fatigue	1.08	0.12	8.78	***
Slowed	0.84	0.14	5.93	***
Agitated	0.02	0.13	0.13	
Age	0.04	0.01	4.07	***
Sex	-0.31	0.25	-1.25	

Table 7-3. *Results of the linear regression analysis (heterogeneity model).*
b, unstandardized regression coefficient; *s.e.*, standard error; *t*, *t*-value; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

7.4.2 Relative importance analysis

The RI estimates of all regressors, representing the allocated individual R^2 contributions of symptoms on impairment, are displayed in Figure 7-1. Different symptoms had drastically different effects on impairment, ranging from RI values of 0.7% *(hypersomnia)* to 20.9% *(sad mood)*. Out of 91 symptom pairs, 76 (83.5%) significantly differed in their RI contributions to impairment (all $p <$ 0.05). RI coefficients within the two compound symptoms *(sleep problems* and *psychomotor problems)* showed differential RI: *early insomnia* (3.6%) was associated with significantly more impairment than *middle insomnia* (0.8%) and *hypersomnia* (0.7%), while *slowed* (8.7%) had a significantly larger RI estimate than *agitated* (2.1%) (all $p < 0.05$).

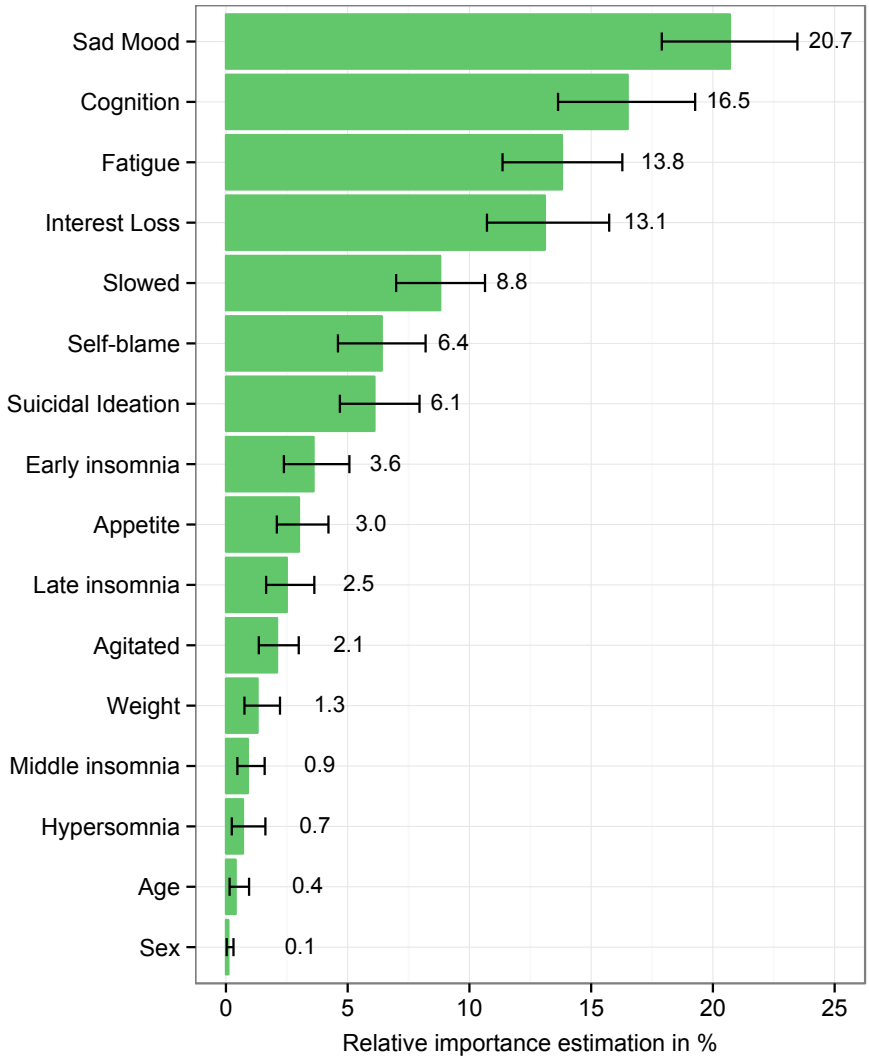

Figure 7-1. *Relative importance coefficients of depressive symptoms on impairment.*
Each value represents the unique shared variance between a symptom and impairment, controlling for age and sex. Estimates are adjusted to sum to 100%.

Are the large differences in the impact of different symptom on disability due to the nature of symptoms, or due to their severity? If severity, then severity differences between symptoms should explain a large proportion of the differences of the RI estimates (i.e. symptoms with high mean values are highly debilitating, whereas symptoms with a low mean are associated with much less impairment). To test this hypothesis we used a linear regression to predict the RI of each of the 14 symptoms by its mean severity. Symptom severity did not reach statistical significance as predictor for symptom RI estimates (F (1,12) = 4.0, p = 0.07). This implies that RI differences are due to symptom nature, and not symptom severity.

7.4.3 Impact of symptoms across impairment domains

Constraining regression weights of symptoms to be equal across the five domains of impairment in model II significantly reduced model fit compared to model I in which symptom contributions were freely estimated ($\chi^2 = 299.8$, $df = 56$, $p < 0.001$). This means that symptoms have differential impacts across impairment domains; these differences between the symptoms-impairment associations across domains are visualized in Figure 7-2. Of the diverse findings, three are especially noteworthy.

1. *Sad mood* and *cognition* were among the four most debilitating symptoms in all domains.
2. *Early insomnia* had comparably strong effects on work impairment, *self-blame* on close relationships, *interest loss* on social activities, and *fatigue* on home management.
3. Compared to other domains, *interest loss* was less impairing for the domain work, *fatigue* for close relationships, *sad mood* for home management, and *cognition* for social activities as well as close relationships.

7.5 Discussion

Overall, individual depressive symptoms have differential effects on impairment, confirming the main hypothesis. Depressed mood, poor concentration, fatigue, and loss of interest explained a large proportion of variance in impairment, whereas weight problems, mid-nocturnal insomnia, and hypersomnia made few unique contributions to impairment.

Subsymptoms within symptom domains had differential effects as well. For instance, psychomotor retardation explained roughly four times as much variance of impairment as psychomotor agitation. These findings highlight not only the importance of considering the nine DSM symptoms individually, but also the importance of considering sub-symptoms within the symptom domains. The three most debilitating symptoms include one affective, one cognitive, and one somatic symptom, suggesting the need to monitor all kinds of depressive symptoms instead of focusing on only one domain or factor score. Furthermore, the two DSM MDD core symptoms, depressed mood and interest loss, made high contributions to explaining impairment, ranking 1 (20.7%) and 4 (13.1%) in general RI estimates. Lastly, although some symptoms were roughly equally debilitating across different domains of impairment, the majority of symptoms varied in their influence across domains.

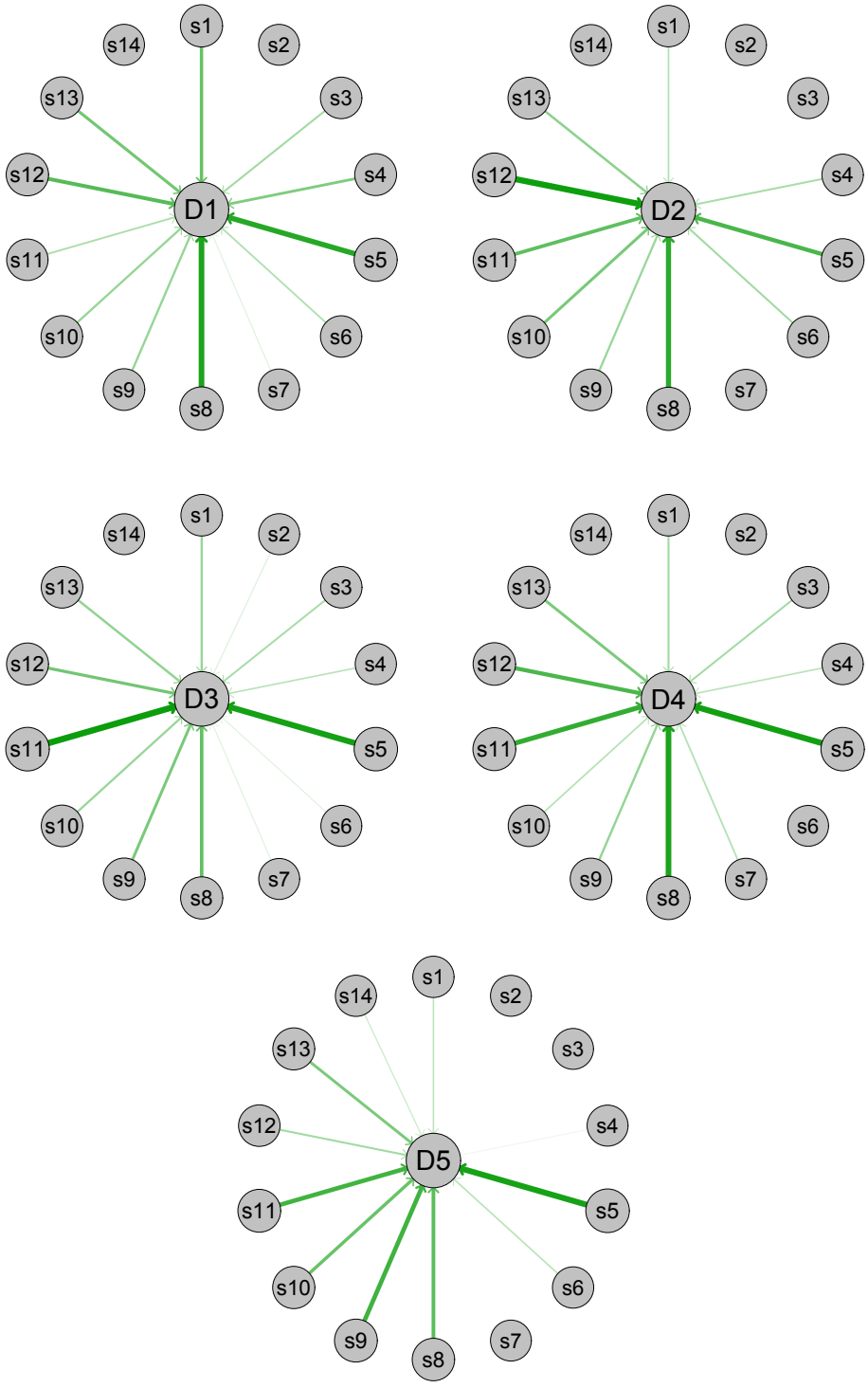

Figure 7-2. *Associations between depressive symptoms and impairment domains.*
The arrows represent standardized regression coefficients of the 14 QIDS-16 depression symptoms (s1-s14) on the five WSAS impairment domains (D1-D5). Thickness of arrows indicates strength of regression weights. D1, work; D2, home management; D3, social activities; D4, private activities; D5, close relationships; s1, early insomnia; s2, middle insomnia; s3, late insomnia; s4, hypersomnia; s5, sad mood; s6, appetite; s7, weight; s8, cognition; s9, self-blame; s10, suicidal ideation; s11, interest loss; s12, fatigue; s13, slowed; s14, agitated.

While prior research has established that symptoms are differentially associated with demographic variables and personality traits (Lux & Kendler, 2010), stressful life events (Cramer, Borsboom, et al., 2011; Keller et al., 2007; Keller & Nesse, 2005, 2006), gene polymorphisms (Jang et al., 2004; Kendler et al., 2013; Myung et al., 2012), and risk factors (Chapter ›5), this study reveals yet another dimension of heterogeneity: symptoms have variable associations with impairment of psychosocial functioning. The broad depression diagnosis not only obscures important differences between patients and lumps individuals suffering from diverse symptoms into the same category – two patients with the same number of depressive symptoms may differ drastically in their functioning levels.

This concealed variability within MDD may substantially contribute to explaining recent disappointing findings such as problems pertaining to depression reliability and the total lack of any confirmed genetic associations (see Chapter ›1.2). The dependent variable in depression studies is usually either a symptom sum-score, or the categorical distinction between depressed and healthy. In both cases, potentially important information about symptoms is lost, and a closer examination of these symptoms is likely to reveal important insights hidden by analyses of sum-scores. In the present study, sleep onset insomnia had comparably severe impact on functioning in the domain of work. It has also been established that MDD treatment is less effective in patients suffering from sleep problems (Dew et al., 1997), that patients with persistent sleep problems are more than twice as likely to remain depressed (Pigeon et al., 2008), and that targeting sleep problems in patients diagnosed with MDD increases overall depression improvement (Lichstein et al., 2000; Rybarczyk et al., 2002). This example elucidates how clinically useful symptom-based approaches can be: they provide detailed information about the nature of problems individuals suffer from, and thus offer the opportunity to improving MDD prevention and treatment.

However, the assessment of individual symptoms does not change the necessity to measure impaired functioning, because the latter may provide a more meaningful construct for actual recovery (Greer et al., 2010; McKnight & Kashdan, 2009). Notwithstanding the fact that measures of functional status are less responsive to treatment than symptom scales and that impairment of functioning often persists after symptoms remit (e.g., Bothwell & Weissman, 1977; Hirschfeld et al., 2000, 2002; Zimmerman et al., 2008), more than 95% of antidepressant trials assessed efficacy via improvements on symptom-scales (McKnight & Kashdan, 2009). This means that only a small minority of reports examined recovery from impaired functioning, which likely leads to an overestimation of treatment success. Especially for the investigation of MDD recovery and efficacy of

antidepressants or psychotherapy, studies urgently need to assess impairment of functioning in addition to individual symptoms of depression.

7.6 Limitations

The results have to be interpreted in the light of five limitations. First, although the impairment scale used in the STAR*D study specifically instructed participants to rate the effects of their depression on functioning, both depressive symptoms and functional impairment were assessed at the same measurement point, so caution about causal interpretations is warranted. Symptoms and impairment potentially reinforce each other and are thus likely to blur, especially in individuals suffering from chronic depression.

Second, while subjects at baseline of STAR*D were not taking antidepressant medication, many participants reported other medical conditions for which prescribed medications might have affected symptom reports.

Third, the bootstrapped CIs for the RI estimates are fairly large for a sample of 3,703 subjects, implying a moderate amount of model uncertainty due to the high number of regressors as well as substantial covariation between them.

Fourth, associations of individual symptoms with impairment may be biased by item wording and should be confirmed with other MDD and impairment scales.

Lastly, differential variability in depressive symptoms is a potential source of biased RI estimates, because heavily skewed symptoms with means close to the minimum and maximum are less likely to demonstrate pronounced statistical relationships. However, symptom means did not significantly predict RI estimates, and even the symptom with the lowest mean of 0.44 (hypersomnia) showed substantial variability ($SD = 0.83$; SD range of all other symptoms excluding hypersomnia: 0.83 to 1.21).

7.7 Acknowledgments

We would like to thank all patients who participated in STAR*D for their kind cooperation.

8 An Alternative Approach to Depression

"At present major depression has become a monolith, with the assumption that the diagnosis can be made merely on the number of depressive symptoms present […]. It may be politically important to utter such simplifications to doctors in general medical setting, but it is a convenient fiction."
– Goldberg, 2011, p. 227

"One does not have to do a deep literature search to encounter the not-altogether-implausible idea that, at the level of the individual person, the symptoms are not effects of a common cause at all; rather they stand in direct causal relations to each other."
– Borsboom, 2008, p. 1101

"I would say, as a rough first stab, that kinds are not simply properties or similarities, but more like congeries of properties held together by laws, i.e. clusters of properties co-occurring because they are lawfully connected."
– Haack, 2003, pp. 131-132

8.1 Key Conclusions

Depressive symptoms vary from each other in important dimensions: they have different risk factors and etiologies, respond differentially to stress, and are differentially associated with biological markers. Symptoms differ from each other in their impact on impairment of psychosocial functioning, interact with each other cross-sectionally in complex patterns, predict other symptoms longitudinally, and have differential predictive effects on depression and other disorders far into the future.

Three main implications can be derived from these findings. First, the studies reviewed above challenge the common cause hypothesis and related assumptions as well as the underlying disease model. While it has been argued that kind essentialism is ill-suited to describe mental disorders in general, it seems especially ill-suited to describe depression. This debate about psychiatry's disease model is more than an eristic discussion amongst philosophers and psychometricians, because important clinical and research practices are based on the assumption of mental disorders as natural kinds. These practices govern everyday decisions of psychologists, psychiatrists, and scientists alike – e.g., by influencing who will be diagnosed with depression, who will receive treatment, how studies are set up, and how data is analyzed. Unless MDD exists as a distinct and clearly demarcated entity similar to infectious diseases described by Koch at the turn of the 19th century, there is little reason to assume that depressive symptoms are caused by depression, that symptoms

are locally independent, or that symptoms are diagnostically interchangeable indicators of a latent condition.

Second, the quest for associated genes, brain correlates, and efficacious treatment is likely to continue to be unsuccessful as long as we uncritically accept consequences of the simplistic and outdated disease model of natural kinds. In contrast, acknowledging the mismatch of theory and data reviewed above is likely going to facilitate our understanding of depression and lead to the development of more accurate models – and ultimately, more efficacious treatment.

Third, results presented in the previous chapters underline the clinical importance of studying individual symptoms; if symptoms are not caused by depression, if they vary from each other in crucial aspects, and if they have direct and causal connections, a new approach is needed that allows for the examination of individual symptoms as well as their lawful associations.

8.2 Previous Approaches to Covert Heterogeneity

The following chapters introduce a *symptom-based approach* to depression and discuss its advantages and limitations. Before moving on to this new approach, it is important to acknowledge that pronounced symptomatic variability across depressed patients has not gone unnoticed in the literature[19]. There have been two common ways to address covert heterogeneity: depression subtypes on the one hand, and methods to extract latent factors or principal components on the other. By and large, however, these approaches have not yielded satisfactory results.

8.2.1 Depression subtypes

Over the last decades, a large number of depression subtypes such as psychotic depression, atypical depression, neurotic depression, endogenous depression, exogenous depression, anxious depression, sociotropic depression, autonomous depression, anaclitic depression, introjective depression, melancholic depression, and hopelessness depression have been proposed (Mcgill, 2011). The main objective of subtyping is to identify homogenous groups of individuals that may be more responsive to specific forms of treatment. A meta-review of 754 reviews published between the years 2000 and 2011 identified 15 commonly mentioned MDD subtypes (Baumeister & Parker, 2012). These subtypes presumably differ from each other in five main aspects: symptom presentation,

[19] Previous subtyping approaches are discussed in this section instead of the introduction, because evidence of covert heterogeneity has been described in detail by now, and because the alternative approach to dealing with heterogeneity is outlined in close proximity (see Chapter ›8.4), allowing for an easier comparison.

etiology, time of onset, gender, and treatment resistance[20]. Unfortunately, subtypes are not clearly demarcated, and no agreement has been reached as to their number or validity (e.g., Baumeister & Parker, 2012; Bech, 2010; Lichtenberg & Belmaker, 2010; Paykel, 2008; Rush, 2007).

A thorough review of all depression subtypes would be a suitable project for a book series, and is beyond the scope of this dissertation. Here, two DSM-5 subtypes are examined that are considered to differ in symptom presentation from "normal" MDD: melancholia and atypical depression.

8.2.1.1 Melancholia

In the DSM-5, melancholia is a MDD-specifier (American Psychiatric Association, 2013, p. 185) characterized by symptoms such as loss of pleasure, lack of reactivity to usually pleasurable stimuli, profound despair, depressive mood especially early in the morning, early-morning awakening, marked psychomotor retardation or agitation, significant anorexia or weight loss, or excessive or inappropriate guilt. Depression with melancholic features was first defined in the DSM-III, and critically discussed at the time of its implementation (Zimmerman & Spitzer, 1989); today, the debate about the validity of melancholia is on-going and unresolved (e.g., Melartin et al., 2004). For instance, it has been argued that the DSM criteria are not specific enough to delineate melancholia from other MDD subtypes or MDD itself, and symptoms such as blunted emotional response, pervasive anhedonia, and reduced libido have been proposed (Parker et al., 2010). Several concepts exist that are closely related to melancholia (e.g., endogenous, endogenomorphic, autonomous, vital, type A, and typical depression), the descriptive validity of melancholia remains questionable, and the consistency of melancholic features across episodes of depression is weak (Baumeister & Parker, 2012; Melartin et al., 2004). It is further problematic that individuals diagnosed with melancholic depression do not seem to differ in their response to antidepressant treatment from non-melancholic patients (Brown, 2007; McGrath et al., 2008), and it is unclear whether *specific* antidepressant treatment is more efficacious in patients with melancholic features. While Perry (1996) reported that TCAs are consistently more effective than SSRIs, another study failed to confirm this notion (Uher et al., 2011). In summary, melancholia remains a contentious subtype that is not clearly demarcated from depression or other subtypes in symptomatology and treatment response.

[20] The DSM-5 notes that depressive disorders differ in their "presumed etiology" (American Psychiatric Association, 2013, p. 155); this expression reflects the ongoing problems of obtaining reliable etiology-subtype associations.

8.2.1.2 Atypical depression

Atypical depression, on the other hand, is characterized by mood reactivity in combination with somatic symptoms such as weight gain, increased appetite, hypersomnia, or leaden paralysis (American Psychiatric Association, 2013, pp. 185). Atypical depression was originally proposed in the 1950s, and early studies found individuals suffering from atypical forms of MDD to be more responsive to monoamine oxidase inhibitors (MAOIs); in 1994, atypical depression was added as MDD specifier to the DSM-IV.

There are various concepts related to atypical depression such as non-endogenous depression, phobic anxiety with secondary depression, vegetative reversal, rejection-sensitivity, and depression with severe chronic pain (see Davidson, 2007). Consistency across and within these categories is low, atypical depression is difficult to distinguish from other MDD specifiers, and validity and reliability of the syndrome remain elusive (Davidson, 2007; Lam & Stewart, 1996; Pae, Tharwani, Marks, Masand, & Patkar, 2009). Furthermore, no modern antidepressant compound with good efficacy has been identified for patients with atypical depression, who respond to TCAs and SSRIs the same way individuals with "typical" depression do (Stewart et al., 2010; Uher et al., 2011); in addition, neither TCAs nor SSRIs have been shown to be superior in the treatment of atypical depression (McGrath et al., 2000; Uher et al., 2011)[21]. These problems have led researchers to conclude that despite 5 decades of research, "the distinct properties of atypical depression remain somewhat mysterious" (Pae et al., 2009, p. 1034).

8.2.2 Depression factors

The second and more data-driven approach to examine symptomatic heterogeneity has been using item-intercorrelations to extract principal components or latent factors from clinical screening instruments such as the BDI, the HAM-D, or the PHQ-9. These factors are often used in subsequent analyses, for example to predict whether life stress is associated with primarily cognitive-affective or somatic depression symptoms (David et al., 2008; Monroe et al., 2001)[22]. The three main problems with approaches aiming to reduce symptom space are summarized as follows.

First, factor solutions differ markedly across and within clinical screening instruments for depression. This means that different instruments come to divergent conclusions regarding the

[21] One study did find fluoxetine (SSRI) to have superior efficacy compared to nortriptyline (TCA) for individuals with atypical MDD; however, the sample of patients with atypical depression only contained 16 individuals (Joyce, Mulder, McKenzie, Luty, & Cloninger, 2004).

[22] Monroe et al. (2001) found life stress to be correlated with only a cognitive-affective factor, while David et al. (2008) identified associations with both depression factors. Both studies used the BDI to assess depressive symptoms, and provide a good example of inconsistencies commonly found in reports of factor-scores.

number and nature of factors in the same dataset, and that factor solutions of the same instrument vary across samples (Brown, Schulberg, & Madonia, 1995; Furukawa et al., 2005; Helmes & Nielson, 1998; Shafer, 2006; Wood et al., 2010). Most surprisingly, the same questionnaire extracts variable factor solutions in subsamples of the same dataset, as demonstrated by Furukawa et al. (2005): "To our own astonishment, randomly selected subgroups from the same population yielded […] different factor solutions, even when we retained 200–500 patients in the sample and employed the same analytic and rotation methods" (pp. 283-284). The CES-D serves as good example for problems pertaining to the reliability of factor solutions; a large number of factor-analytical papers have been published, and while some confirm the originally proposed 4-factor structure by Radloff (1977) (e.g., Blazer, Landerman, Hays, Simonsick, & Saunders, 1998; Iwata & Roberts, 1996), solutions with one to three factors as well as higher-order factor structures have been identified (e.g., Helmes & Nielson, 1998; Lee et al., 2008; Morin & Ninot, 2011; Wood et al., 2010).

Second, different approaches of extracting factors such as principal component analysis and confirmatory factor analysis (CFA) can lead to different results in the same dataset (e.g., Widaman, 1993).

Third, screening instruments have failed to demonstrate consistent *measurement invariance* (also called *measurement equivalence*). Tests of measurement invariance are conducted to examine whether a given questionnaire assesses the same construct in different populations (i.e. whether a test is invariant across samples), for example after translating an instrument into another language (Schmitt & Kuljanin, 2008). To do so, increasingly constrained CFAs are compared. There are various levels of measurement invariance; simply put, if basic forms of invariance cannot be established, a statistical comparison of different groups is meaningless because important model parameters such as the number of factors or factor loadings differ across these groups[23]. For clinical screening instruments of depression, measurement invariance violations have consistently been demonstrated, and factor solutions are not invariant for important variables such as ethnicity (e.g., Crockett et al., 2005), sex (e.g., Baas et al., 2011), or age (e.g., Williams et al., 2007). This means that sum-scores as well as factor-scores often assess different constructs in different populations.

8.3 The Nosological Predicament: Splitting or Lumping

Despite the problems described above, factor scores and depression subtypes are by no means useless concepts; on the contrary, building etiologically and symptomatologically homogenous subgroups that show superior response to specific treatment should be one of the most important

[23] For a detailed explanation of measurement invariance analysis in depression research, see Baas et al. (2011).

goals for a highly heterogeneous disorder such as MDD. A commendable example of using exploratory methods to reduce heterogeneity with promising results for treatment are two connected studies in which a depression factor consisting of the symptoms 'loss of interest', 'diminished activity', and 'inability to make decisions' predicted poor antidepressant response (Uher et al., 2008, 2012). Overall, however, success has been limited, and due to the potential problems of subtypes and methods that aim to reduce symptom space the subsequent focus is on individual MDD symptoms.

An important problem in nosology is the question of splitting and lumping: how far apart do two disorders have to be in a multidimensional diagnostic space to classify them as two separate diseases? While psychiatry has had a general bias towards lumping, the DSM-5 adopted an even more pronounced stance for which the *bereavement exclusion* is a good example. If an individual recently suffered the loss of a loved one and shows symptoms such as sad mood, insomnia, and loss of interest, the DSM-IV provided clinicians with the option to understand these symptoms as grief-related response to a loss, in which case a diagnosis of clinical depression was not necessarily indicated; grief was conceptualized as normal response to a terrible situation, in contrast to the clinical condition MDD. This position was taken because bereavement can be uncomplicated, short-lived, benign, impairs functioning less than clinical depression, and can remit without specific treatment (Friedman, 2013; The Lancet, 2012; Wakefield, 1997)[24]. In the DSM-5, however, the bereavement exclusion was removed, lumping more individuals into the already highly heterogeneous category MDD.

I believe that we may have erred on the side of lumping for a prolonged time now, with unsatisfactory results, and propose an approach in the following sections that may lean towards erring on the side of splitting. Once we have understood core units of depression that have received so little attention thus far, we can move on to lumping once more. However, a detailed understanding of symptoms will be necessary for creating valid and reliable diagnostic subcategories.

8.4 A Symptom-Based Approach to Depression

The idea that symptoms may differ from each other in important aspects is not new. Roughly a decade ago, Hasler et al. (2004) argued that individual depression symptoms like low mood or sleep problems may represent separate *endophenotypes*, potentially differing from each other in aspects like

[24] See also Zisook et al. (2012) who argue in favor of removing the bereavement exclusion, as well as DSM-5 p. 811 for a list of reasons for the removal.

pathophysiology, risk factors, brain correlates, and associated genes[25]. Even years before that, Persons suggested isolating single elements of pathology for study (1986), and Costello recommended the investigation of individual symptoms of depression (1993). In spite of these early calls, a systematic review of symptomatic heterogeneity as well as several related studies specifically investigating different dimensions of covert heterogeneity had not been conducted prior to this dissertation.

The focus on symptoms, however, is only one of three necessary components to a new symptom-based approach. Additionally, new models have to accommodate the facts that symptoms are highly comorbid, and that symptoms possess individual causal power. This means that they need to be integrated into a framework that allows symptoms to be associated with each other beyond a latent variable explaining their covariance.

8.4.1 The network view of depression

Very recently, Denny Borsboom and colleagues at University of Amsterdam put forward a network approach to psychopathological symptoms (for an overview, see Borsboom & Cramer, 2013), integrating the focus on individual symptoms as well as the relations between symptoms, with a strong focus on the latter. The model fits the concept of symptoms as endophenotypes outlined above, complements the research conducted in Chapters ›5, ›6, and ›7, and is in line with the symptom-based evidence reviewed in Chapter ›4. Conceptually, the approach is straightforward: psychopathological symptoms are autonomous causal entities in complex dynamical networks. Depression is not understood as a latent disease; instead, it is constituted by causal connections of symptoms to each other. In networks, symptoms do not represent indicators of an underlying disorder – they *are* the disorder. This perspective deals elegantly with many of the problems described in Chapters ›3 and ›4 and moves away from depression as distinct natural kind.

Figure 8-1 visualizes differences between a traditional reflective latent variable model and a cross-sectional symptom network. In this hypothetical network that I constructed merely for purposes of explanation and that is not based on real data, especially symptoms s3, s10, s11, and s12 are closely interconnected and central, while others such as s6 or s13 have only few weak connections; s2 is altogether uncorrelated. It is important to note that this particular network is cross-sectional, and links between symptoms (e.g., between s10 and s11) reflect bidirectional associations, not unidirectional causality.

[25] Endophenotypes, sometimes also referred to as *intermediate phenotypes* or *biological markers*, are intermediate processes that form causal links between genes and overt behavior (Cannon & Keller, 2006; Gottesman & Gould, 2003).

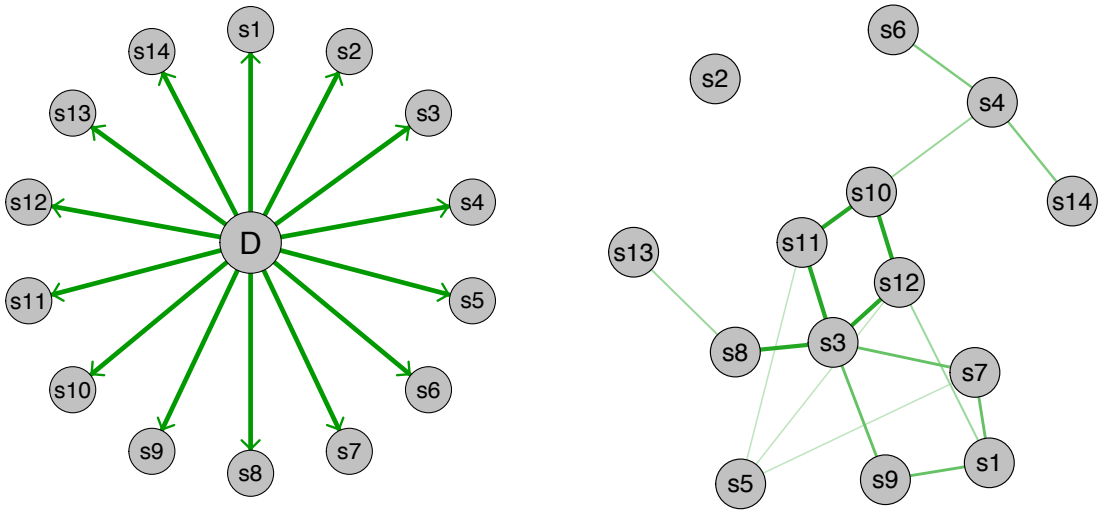

Figure 8-1. *Comparison of a reflective latent variable model and a network model.*
Left side: reflective latent variable model in which the underlying disease depression (D) is the common cause for the depression symptoms s1-14. Right side: network model in which depression consists of the associations between symptoms. The green lines in the network represent correlations between symptoms, and line-thickness indicates correlation-strength. The network configuration is derived from an algorithm reflecting the strength of associations: highly correlated symptoms are displayed in the center, symptoms with weak correlations in the periphery. Both plots were constructed using the R-package QGRAPH (Epskamp et al., 2012).

8.4.2 Homeostatic property clusters

The statistical model of symptom networks is derived from the philosophical concept of *homeostatic property clusters* (HPCs)[26]. In a school of thought that has been termed *scientific realism*, Richard Boyd introduced this idea to describe biological species (Boyd, 1999; Richard Boyd, 1991; Wilson et al., 2007), contrasting kind essentialism. According to Boyd, a species is not a distinct natural kind with clearly demarcated boundaries, because variation and heterogeneity within a species is not a deviation from the true essence of a biological kind, but part of what it *is* to be a member of those kinds (Ereshefsky, 2007; Wilson et al., 2007).

Generally, Boyd argues that properties of a specific family of properties (e.g., F_1) are contingently clustered in nature and appear together reasonably often because the presence of one property tends to favor the presence of another (Boyd, 1999). We are inclined to describe such HPCs as natural kinds (Zachar & Bartlett, 2002), but since relationships between properties are often probabilistic and not deterministic, imperfect aggregations of properties exist, meaning that some thing may show most but not all properties of F_1 (this may, in fact, be the rule rather than

[26] HPCs can also be referred to as *mechanistic property clusters* (MPCs); Kendler et al. (2010) suggested to use the latter name because the term "homeostatic" may be misleading, given its meaning in the physiological literature.

the exception). From this vantage point, species are classes that share related features (e.g., genetic, behavioral, and physiological) due to the existence of a multitude of underlying causal mechanisms lawfully connecting these properties, and the large majority of individuals of a specific species are more closely clustered on a multidimensional space of properties than individuals of other species. However, different clusters overlap and are not encapsulated or distinct.

Psychopathological symptom networks utilize this view and understand mental disorders as clusters of properties (symptoms) that are related due to underlying causal mechanisms (e.g., direct causal links between symptoms). This stresses the importance of investigating these mechanisms, and promises substantial advances that are discussed below. The model is still in its infancy, and potential problems are highlighted as well.

8.4.3 Advantages of a symptom-based approach

8.4.3.1 Focus on individual symptoms

Of the numerous benefits, maybe the most important one is that network models shift the focus from problematic sum-scores to individual symptoms and their causal links. They acknowledge symptom-importance and encourage their analysis instead of using aggregate scores or latent factors. The most unconventional hypothesis is to abolish depression as latent variable altogether. This is similar to the way HPCs treat categories: as names that may be useful in describing families of properties, but it is the properties that hold causal power, not the categories themselves. Depression from this perspective is not an underlying disease in a reflective sense that causes symptoms; instead, symptoms are considered separate entities that deserve attention as part of the disease, not as its consequences.

8.4.3.2 Centrality

The network approach does not require problematic assumptions of reflective latent variable models such as the common cause hypothesis or the notion of local independence. Symptoms are different entities that directly interact with each other across time due to underlying mechanisms, and not the product of one common cause. The covariance of symptoms does not need to be explained by a latent variable, because depression is understood to be a complex constellation of a large number of variables instead of one latent structure. This means that a network perspective does not presume symptoms to be equivalent or exchangeable; while indicators are by definition equally central in a reflective latent variable model (see Cramer et al., 2010), networks allow for the assessment of symptom centrality as model parameter. Centrality in this sense is a measure of interconnectedness of a given symptom in a network of symptoms (Boccaletti, Latora, Moreno,

Chavez, & Hwang, 2006; Borsboom & Cramer, 2013): a very central symptom has a large number of connections with other symptoms (e.g., s3 in Figure 8-1), while an isolated symptom with few connections exhibits low centrality (e.g., s6 in Figure 8-1). As described in Chapter ›4.3, there is evidence that the treatment of specific symptoms is an effective intervention strategy, and different schools of psychotherapy have long embraced this notion. Research could potentially inform clinicians which symptoms are most central in psychopathological networks, and specifically targeting highly central symptoms and thus removing them and their causal power may prove to be an efficient strategy.

Interestingly, the notion of centrality corresponds to the way clinicians intuitively think about psychiatric disorders – as clusters of causally related variables. In several experiments, Kim and Ahn (2002) demonstrated that when clinicians are confronted with symptom reports from mental disorders such as MDD, anorexia nervosa, schizophrenia, antisocial personality disorder, and specific phobia, patients with causally central symptoms are more likely to be diagnosed with a given disorder, causally central symptoms are judged to be more typical symptoms of a given disorder, and clinicians recall causally central symptoms with greater accuracy than peripheral symptoms. The authors concluded that the causal thinking exhibited by clinicians is in stark contrast to the atheoretical DSM approach of un-weighted symptom lists.

8.4.3.3 External factors as variables to group networks

The network approach allows for the integration of external factors such as personality traits or life events into models[27]. These variables can be used as grouping factors to create more homogenous networks; for example, the network in Figure 8-1 could be estimated separately for men and women, and then compared. In a recent study of the National Comorbidity Survey Replication (NCS-R) – a nationally representative survey of the incidence and prevalence of mental disorders among English-speaking adults of the U.S – depressed men reported higher rates of anger, substance abuse, and risk taking, which were interpreted by the authors as specifically male depression symptoms (Martin, Neighbors, & Griffith, 2013). While the paper did not explicitly understand or analyze these symptoms as part of an interconnected network, the data offer the great opportunity to explore male and female symptom networks: anger may not only be more prevalent, but also more *central* in male networks.

---

[27] The terms *external factors* and *external variables* in this and the following sections refer to potentially relevant network variables that are not depression symptoms, e.g., personality traits, demographic variables, life events, or genetic dispositions.

In a cross-sectional study conducted by Cramer et al. (2011) that explored group differences of networks, the configuration of relationships between depressive symptoms (i.e. the correlation matrices of symptoms) varied across four groups of subjects that experienced four different adverse life events. For instance, the symptoms depressed mood and suicidal thoughts were strongly associated in individuals that had previously experienced health problems, while the link of these symptoms was much weaker in a group that had suffered a romantic loss. The study underlines the importance of the impact of external variables on symptom networks, although it remains unclear what particular mechanisms govern the development of specific symptom associations in response to particular forms of life stress.

8.4.3.4 Complex models and feedback cycles

Highly complex models are possible as well. While to my knowledge no such network has hitherto been constructed, external variables could be integrated into networks themselves, for instance as moderators between connected symptoms. Certain risk factors may increase the probability to develop a particular symptom in response to another symptom. A good example is the pathway from insomnia to fatigue for which substantial inter-individual variation has been demonstrated (Achermann, 2004).

External variables could also increase the probability to develop emergent symptoms; specific life events, for instance, could lower the threshold to develop sleep problems, subsequently leading to other MDD symptoms. This, in turn, might depend on vulnerability factors for this particular link, e.g., a high neuroticism score, or a specific genetic disposition.

It is also possible that symptoms influence symptom-to-symptom links (e.g., insomnia might affect concentration problems via fatigue), or that symptoms are integrated in feedback loops. Clinicians have long understood the importance of vicious circles in the development and preservation of psychological problems; for example, many patients suffering from anxiety disorders avoid feared stimuli such as particular animals, objects, or places. This reinforces aversive emotions because the escape of a potentially devastating situation is experienced as rewarding: "Because I did not take the elevator, nothing happened to me. Had I taken the elevator, I would have died". Fear and avoidance – symptoms of several anxiety disorders – are coupled in a feedback loop, and exposure therapy in which individuals are slowly habituated to the feared stimulus is often successful because it helps to break this self-sustaining circle of symptoms (Emmelkamp, 2003). Whereas vicious cycles are well-established in the clinical literature, traditional latent variable models cannot account for symptoms that increase in severity due to other symptoms due to the assumption of local independence (see Chapter ›3.3). Nonetheless, it is plausible that symptom

networks can be self-sustaining, and even if a clinician successfully treats a specific symptom such as concentration problems in a patient diagnosed with MDD, it is likely to reappear *due to* other persisting and causally related symptoms such as insomnia and fatigue.

This makes residual symptoms a potentially interesting unit of study for networks: depending on the nature of residual symptoms, it is likely that very specific other symptoms are activated. For instance, Dombrovksi et al. (2007) showed that residual anxiety and sleep problems independently predicted earlier MDD recurrence in patients who had previously remitted from late-life depression. Understanding the causal mechanisms that underlie residual symptoms may lead to more efficient strategies to specifically target populations at high risk for recurrence.

8.4.3.5 Inter-individual variation

If one symptom causally leads to a second and then to subsequent symptoms, possibly in interaction with life events, stress, and other personal variables, the term "depressed" does not describe a clearly demarcated state with a distinct pathophysiology, but instead different situations with pronounced inter-individual variability. The network approach offers a more realistic view of the large variety of symptoms depressed patients suffer from, and elucidates that a clinically depressed population does not represent a homogeneous group of individuals suffering from one distinct disorder. Symptom associations likely vary across individuals, and certain symptom links will be much more closely associated in some individuals, potentially depending on genetic dispositions, coping behavior, or habits such as drinking or doing sports. This means that depression networks are explicitly etiologically heterogeneous, and there are numerous ways in which external factors such as stress and lack of social support, or internal factors such as a predisposition towards low mood can lead to the development of initial and emergent depressive symptoms. Most importantly, the approach not only acknowledges that two patients with the same number of symptoms may suffer from different symptoms – it is possible (and plausible) that patients with the same symptoms "got there" via two very different psychopathological and pathophysiological pathways.

To give one example, Durmes and Dinges (2005) established substantial inter-individual variance in sleep deprivation in a review of neurocognitive responses to insomnia. They concluded that this might be due to "traitlike differential vulnerability to impairment from sleep loss, for which neurobiological correlates have yet to be discovered" (p. 123). This stresses the importance of modeling inter-individual differences in symptom presentation, potentially offering the opportunity to facilitate our understanding of the complex mechanisms governing the biological roots of depression.

8.4.3.6 Missing heritability

Inter-individual differences of direct causal paths between symptoms may offer an answer for a pervasive problem in psychiatry: the so-called "mystery of missing heritability" (Johnson, Penke, & Spinath, 2011; Zuk et al., 2012). Notwithstanding the fact that depression is a moderately heritable disorder, with heritability estimates ranging between 37% and 60% (Boomsma, Busjahn, & Peltonen, 2002; Kendler, Gatz, Gardner, & Pedersen, 2006; Sullivan, Neale, & Kendler, 2000), decades of gene-hunting have failed to provide replicated genome-wide results of relevant genes (see Chapter ›1.2.4). Symptom-networks offer an explanation for this paradoxical finding: current approaches aimed at identifying genetic processes underlying MDD employ sum-scores or binary indicators (depressed vs. healthy) which can only capture genetic variance that all depression symptoms of a given questionnaire share (Cramer, Kendler, & Borsboom, 2011). Common variance among symptoms, however, may be low, and the lack of findings is not surprising from this perspective. Genetic predispositions may mediate specific causal structures between symptoms, and in some individuals, the activation of a few symptoms may reliably lead to a plethora of other symptoms. Analyses of sum-scores are unable to detect such effects, and investigating genetic variability of direct symptom pathways may substantially contribute to uncovering missing heritability.

8.4.3.7 Comorbidity

Understanding different psychiatric diagnoses as families with related properties elegantly deals with the problem of comorbidity that has been documented in large community studies, for instance between generalized anxiety disorder (GAD) and depression (Kessler et al., 2005; Mineka et al., 1998): roughly half of the patients diagnosed with MDD also have a diagnosis of GAD and vice versa[28]. Nonetheless, MDD and GAD are commonly understood to be separate disease kinds, and from a traditional perspective, co-occurrence implies that an individual suffers from two distinct disorders. This is explained by a general susceptibility towards negative affect, or by shared genes that predispose for both disorders (Barlow, Allen, & Choate, 2004; Mineka et al., 1998).

[28] While GAD is used as example of a diagnosis highly comorbid with MDD, the results are similar for other disorders such as PTSD: there is substantial symptom overlap for PTSD and MDD (e.g., Rosen et al., 2008), and comorbidity rates in population-based surveys range from 62% to 92% (Keane, Brief, Pratt, & Miller, 2007; Perkonigg, Kessler, Storz, & Wittchen, 2000). A subset of PTSD symptoms is closely related to symptoms of anxiety and mood disorders, and a recent study demonstrated that participants with MDD consistently reported similar sum-scores on various PTSD screening instruments compared to individuals with PTSD diagnosis (Gros, Price, Magruder, & Frueh, 2012).

From the vantage point of property clusters, however, it is very likely that individuals in F_1 will often be found in F_2 in case both families share defining properties. This is the case for MDD and GAD: there is considerable symptom overlap between the diagnoses (insomnia, fatigue, and concentration problems), and both property clusters thus overlap in a multidimensional property space. In psychopathological network models, such symptoms that are part of several diagnostic categories can be collapsed into one node; e.g., fatigue will be present only once in a network of MDD and GAD symptoms. These bridge symptoms may play a crucial role for the development of comorbidities (cf. Cramer et al., 2010). An individual reporting only depressive symptoms in the beginning could potentially develop several bridge symptoms, eventually leading to comorbid GAD: sad mood (MDD) → loss of interest (MDD) → sleep problems (MDD/GAD) → fatigue (MDD/GAD) → concentration problems → (MDD/GAD) → irritability (GAD). Symptoms activate other symptoms, and criteria for two separate diagnoses are met. This, of course, could go the other way around: while one individual may develop GAD symptoms in response to MDD symptoms, the direction of causation may be inverted for another patient.

In a community-based sample of 1,014 youths between the ages of 13 and 16, anxiety disorders were predictive of subsequent insomnia onset (OR = 3.5), whereas insomnia predicted depression onset (OR = 3.8) (Johnson, Roth, & Breslau, 2006). While individual symptoms were not analyzed in this report, it is plausible that insomnia may connect GAD and MDD, and that the majority of individuals develop MDD after GAD. Insomnia as a bridge symptom may also offer a potential explanation as to why women in general seem to be at a greater risk to develop not only primary insomnia, but also GAD and MDD (Krystal, 2003).

However, not only psychiatric diagnoses share symptoms with the diagnostic criteria of MDD: medically ill patients often report symptoms such as fatigue and insomnia, which could artificially increase depression rates in such populations (Zimmerman, Chelminski, McGlinchey, & Young, 2006). Conceptualizing both medical and psychiatric symptoms in potentially closely entangled symptom-networks – and acknowledging the fact that various symptoms appear in many different disorders – may substantially improve the field's grasp of comorbid conditions by allowing the investigation of causal pathways into comorbidities.

8.4.3.8 Susan and Paul

Susan and Paul are two patients that recently walked into the office of a clinician. Susan is 27 years old, and started her residency training as a surgeon few months ago. Susan is generally not a good sleeper, slightly anxious, tends to worry a lot, and her father has had several episodes of MDD. Although Susan's medical education has been demanding so far, she is highly intelligent and did

very well in school and university. For the year of her internship she has to move to a different city several thousand miles from home, and can see her fiancé only once a month. Internship soon confronts her with a 90-hour week, and while most of her colleagues fall asleep as soon as they lie down on a flat surface, Susan suffers from insomnia. This leads to pronounced levels of fatigue as well as psychomotor problems, and she commits errors in the hospital. Susan begins to worry about her career as a surgeon, further increasing her sleep problems, and soon she is trapped in a vicious cycle of worry, insomnia, fatigue, and errors at work. She feels terrible about her mistakes and soon questions herself every time she performs a medical procedure, and eventually starts seeing a psychiatrist.

Paul is 45 years old, and usually falls asleep within minutes and wakes up after 7 hours, feeling well and refreshed. He has not had previous episodes of MDD and never been to a clinician before. Paul is very active and works out several times a week, but has a bike accident and is in serious pain for 3 consecutive months. The pain prevents him from sleeping well, and also from doing any kind of sports, which turns out to be very important for his well-being. Pain and insomnia subsequently lead to low levels of fatigue, loss of energy, and negative mood; Paul's wife has problems recognizing her husband who is usually very kind and calm, and the two argue a lot, further exacerbating existing symptoms and leading to emergent symptoms such as psychomotor retardation and concentration problems. Paul, like Susan, starts consulting a clinician.

Both cases demonstrate how different pathways into depression are, and the network perspective offers a new approach to investigate these pathways that are incompatible with the common cause hypothesis and various related assumptions. Depression is a network of comorbid symptoms that are lawfully connected, and symptom associations for both Susan and Paul depend on a plethora of other variables such as genotype, environment, personality, life history, and social support.

8.4.4 Limitations of a symptom-based approach

While there are substantial benefits to modeling symptoms in a network of associated variables, I see difficulties and room for improvement in the following areas.

8.4.4.1 Boundaries between disorders

Symptoms like insomnia are currently understood to be indicators for different and often comorbid disorders. A symptom-based approach no longer distinguishes between 'GAD insomnia' and 'MDD insomnia', because sleep problems are not consequences of either disorder. Insomnia is likely an important and central symptom in both MDD and GAD, making the decision where to

exactly separate the interconnected disease networks a very difficult one because sharp factual boundaries – in the sense of distinct natural entities – may simply not exist.

The network approach acknowledges (and in fact highlights) these fuzzy boundaries, but has so far not provided an alternative nosological approach. Diagnostic categories, however, are needed to further study mental disorders and provide treatment to individuals suffering from various conditions. A definition of depression is what makes studies from different countries and cultures comparable, is required to examine efficacy of drugs and psychotherapy, and the provision of health-care is hard to imagine without a classification system. From the perspective of HPCs, a possible solution may be to use a heuristic value that informs clinicians how well as person fits into (and deviates from) various psychopathological families, i.e. how many properties of each family are fulfilled[29]. Overall, further research will be required to address this substantial problem of classification.

8.4.4.2 Complexity

Networks are complex and dynamic processes with a high degree of individualization; this complexity is an intrinsic property of many psychological constructs, including mental disorders (Mcgrath, 2005). While estimating different networks of potential causations between symptoms and external variables for ten different patients diagnosed with MDD may provide clinically relevant and practical insights for these ten individuals, the approach is impractical from a more general epidemiological perspective.

Psychiatry's main goal is to develop more efficient interventions. Progress to date has been slow for depression, and identifying MDD subtypes that are responsive to specific antidepressants or particular forms of psychotherapy is the next logical step. Hence, we are faced with the question how to deal with the massively heterogeneous symptom networks, and what cut-offs can be used to determine meaningful differences between networks. As discussed previously, Cramer et al. (2011) demonstrated that cross-sectional symptom networks of individuals that had previously experienced four adverse life events differed from each other; they did so by estimating a homogeneity model in which the correlations between symptoms were constrained to be equal across the four life event categories, and tested this model against a heterogeneity model in which symptom-correlations were freely estimated across categories. While the statistical approach is sound, the overall information that there are statistical differences between networks conceals how different

[29] Contrasting an essentialist perspective, these heuristic coefficients would be continuous and dimensional rather than categorical, and thresholds probabilistic rather than deterministic.

the symptom networks exactly are, what symptom paths are responsible for the symptom configurations to vary across life events, whether networks are meaningfully different from each other in a clinical sense, and whether they are different enough to understand symptom responses to particular life events as separate MDD subtypes.

This is not meant as criticism – the study of Cramer et al. (2011) provides new and important insights. However, the paper allows for the opportunity to understand what future research will have to entail in order to answer the bigger questions psychiatry is faced with. The ultimate goal of a symptom-based approach is to establish several useful sub-categories that capture a majority of patients within the super-diagnosis depression. While networks are an excellent starting point, one of the most important next steps is the development of statistical approaches to somehow cluster symptom networks.

8.4.4.3 Grouping variables

External variables such as distinct life events (Cramer et al., 2011) have been used as grouping factors for symptom networks, showing that symptom associations vary across different life events. When introducing a sufficiently large number of grouping variables, however, the approach soon becomes impractical due to combinatory explosion. For example, the question how nine MDD symptoms interact with each other regarding sex, history of depression, and four distinct life events would necessitate the estimation and comparison of 16 networks[30] with 36 intercorrelations[31] per network. These correlation networks could be compared using an omnibus test as was done by Cramer et al. (2011). More specifically, one could examine whether the symptom inter-correlations of network I (e.g., sex = male; history of depression = yes; life event = romantic loss) and network II (e.g., sex = female; history of depression = yes; life event = romantic loss) differ from each other. With 16 networks, there are 120 comparisons of two networks that would need to be performed[32].

With a higher number of external variables, it may thus be preferable to develop models in which such external variables can be included directly into the networks. As discussed in Chapter ›8.4.3.4, these variables may moderate particular links (e.g., between symptoms, or between risk factor and symptoms), and estimating effects of external variables in symptom networks would allow for the development of specific prevention and intervention strategies for particular populations.

---

[30] 2 (sex) x 2 (history of depression) x 4 (life events) = 16.

[31] 9 choose 2 (unordered with no repeats): 9! / (2! x (9-2)!) = 36.

[32] 16 choose 2 (unordered with no repeats): 16! / (2! x (16-2)!) = 120.

9 Implications of Covert Heterogeneity

> *"The most widely used instruments in clinical settings have generally failed to provide clear documentation of the symptoms experienced by individuals and instead typically have offered only global indices of depression. Assessment tools [...] need to deal effectively with the heterogeneity of depression."*
> *– Strategic plan for mood disorder research, NIMH, 2003, p. 93*

Depression does not have an essence, and psychiatry has been unable to identify pathognomonic depression symptoms or other markers that allow for a reliable diagnosis of MDD. Depression is in many ways different from infectious diseases, and does not fit into the framework of natural kinds.

What are the implications of the pronounced heterogeneity of depressive symptoms, and what should future studies aiming to uncover important symptom-based information take into consideration? The following sections focus on the improvement of quality and quantity of symptom assessment, the development of methods to cluster networks, the inclusion of external variables such as life stress into psychopathological network models, and the longitudinal study of networks. Furthermore, symptom-based datasets (e.g., from clinical trials) should be made available to researchers for the purpose of reanalysis, nomenclature should change to allow for easier identification of symptom-based research, and focused work on the concept of sum-scores is needed. Finally, I discuss how well these implications generalize to other mental disorders such as PTSD or GAD.

9.1 Symptom Assessment

First and foremost, the field needs to obtain a good picture of the symptoms that individuals classified as depressed suffer from. However, since classifications are *defined* by the presence of symptoms, we find ourselves in a tautological quandary.

My suggestion to address this problem is to measure many more symptoms than we currently do in order to provide an empirical baseline; this will allow us to answer the question which symptoms are most central in psychopathological networks. Curiously enough, this very basic information is unavailable at present; the following section introduces current problems pertaining to symptom assessment and concludes with implications for future studies.

9.1.1 Problems with symptom assessment

There are six main problems with the current assessment of depressive symptoms. First, today's MDD criterion symptoms were determined largely by clinical consensus instead of empirical evidence (Kendler & Zachar, 2008; Lux & Kendler, 2010; Zimmerman, McGlinchey, Young, & Chelminski, 2006a). One of the first proposed sets of symptoms goes back to the 1957 report by Cassidy et al. (1957) who described clinical features of manic-depressive disorders. The list was reworked later by Feighner et al. (1972), but no data were published to support these changes. The DSM-5 symptoms closely resemble the ones proposed over 40 years ago; this is baffling, considering the numerous critical calls for a psychometric (re-)evaluation of depression and its symptoms (e.g., Andrews et al., 2007; Lux & Kendler, 2010; McGlinchey et al., 2006; Mitchell et al., 2009; Zimmerman et al., 2006a, 2006b, 2006c).

Second, different scales that presumably assess the same latent disorder do so by measuring a broad range of different depression symptoms (Shafer, 2006), and of the four widely used scales investigated by Shafer, not even one covers all DSM criterion symptoms of depression[33].

Third, the instrument that is considered to be the gold standard for depression assessment – the HAM-D – was developed over 5 decades ago, and a review of 70 studies demonstrated that it does not possess satisfactory psychometric properties (Bagby, Ryder, Schuller, & Marshall, 2004): many HAM-D items measure depression severity poorly, several items have poor inter-rater and re-test reliability, the response format is not optimal, and content validity is poor. Nonetheless, the HAM-D is used as main outcome measure in the majority of clinical trials that assess antidepressant efficacy, and often used to assess whether individuals qualify for depression trials[34]. Furthermore, when new scales are developed, sum-scores on these instruments are compared to sum-scores on the HAM-D to "validate" the new instrument.

Fourth, the idea of *internal consistency* in scale development is problematic for depression instruments. Coefficients such as Cronbach's alpha are commonly used to evaluate the psychometric quality of a given scale: the higher the coefficient, the better the instrument. Internal consistency is based on the intercorrelations of items, and if intercorrelations are high, it can be concluded that items measure the same underlying construct. This approach usually leads to dropping problematic items that do not show substantial correlations with other items, a process that iteratively improves

[33] The study investigated the BDI, the HAM-D, the Center for Epidemiological Studies Depression Scale (CES-D), and the Zung Self-Rating Depression Scale (SDS).

[34] In the NIH-sponsored STAR*D dataset used in Study 3 of this paper (Chapter ›7), a sum-score of ≥ 14 on the HAM-D was necessary for participation.

internal consistency. While dropping items with low correlations may be sensible when measuring a latent construct that is the common cause of its symptoms, it is highly problematic for depression, given the evidence presented in this dissertation. If only one in five depressed patients suffers from a specific symptom (e.g., anger), the correlations with other symptoms in the whole population may be comparably low. To improve internal consistency, the symptom would commonly be excluded from the scale, although it may be tremendously debilitating and central for those individuals suffering from it, resulting in loss of important data.

Fifth, there is no single pathognomonic depression symptom that indicates depression beyond reasonable doubt. This is different from other disorders: sudden chest pain that radiates into the left arm or neck, for instance, is a symptom that is common in individuals with heart attacks, and is only common in individuals with heart attacks. The same is true for Koplik's spots: individuals with measles often exhibit them, but they never appear in individuals without measles. Headaches, on the other hand, are common in many different disorders, have a large number of comorbid symptoms, and subsequently low sensitivity and specificity; they do not seem especially useful to determine which particular illness is present. Unfortunately, the majority of the DSM-5 criterion symptoms for depression are, at least from a diagnostic perspective, much more comparable to headaches than to sudden chest pain or Koplik's spots.

Lastly, there are marked differences between standardized screening instruments for depression in the classification of depressed patients into severity groups (Zimmerman et al., 2012); this means that the *nature* of the scale chosen for a particular study biases which participants qualify for clinical trials, and how many individuals achieve remission in these trials.

9.1.2 Implications for symptom assessment

An essential next step towards illuminating covert heterogeneity is the assessment of a larger number of depression symptoms that may not be as prevalent as insomnia or depressed mood, but may nonetheless causally contribute in symptom networks. Potentially interesting symptoms are anxiety and anger/irritability, because they are highly prevalent in depressed patients, and seem to be associated with worse clinical outcomes. In a study of 2,876 outpatients diagnosed with MDD, 53.2% reported anxiety symptoms (Fava et al., 2008). In this anxious group, remission of depression was less likely and also took longer. Furthermore, a recent study demonstrated that 54.4% of 536 depressed individuals exhibited overt irritability and anger (Judd et al., 2013), and that these symptoms were markers of a more severe, chronic, and complex depression.

In addition to assessing more symptoms, subsymptoms need to be distinguished from each other properly. When looking at the bi-directional DSM-5 criterion symptom 'insomnia or hypersomnia' that is commonly treated as one item in depression questionnaires like the PHQ-9, it is evident that insomnia potentially plays a very different causal role than hypersomnia, and subsuming both into one category will hamper the progress of uncovering covert heterogeneity. While this may ultimately lead to the assessment of many dozen symptoms, I suggest that it is better at present to err on the side of splitting than to continue to err on the side of lumping. Splitting allows data-driven lumping, while lumping leads to loss of potentially relevant data. Once insomnia and hypersomnia have been assessed with one question, it is impossible to disentangle which patient slept too much and which slept too little; however, if two different items insomnia and hypersomnia should turn out to be highly correlated, they can still be grouped together in subsequent analyses.

Sleep problems can be explored even further, and questions regarding nightmares may be worth including in future depression questionnaires, since individuals suffering from nightmares showed a 5-fold increase in risk for high suicidality, even after adjusting for psychiatric diagnosis and psychiatric symptom intensity (Sjöström et al., 2007). Moreover, the problem of lumping subsymptoms into symptom domains does not only pertain to sleep problems. Fatigue is another example, and Ferentinos et al. (2009) describe the problem well when they state: "sleepiness and fatigue are conceptually distinct constructs: insomnia causes fatigue, while sleep apnea and narcolepsy cause mostly daytime sleepiness; fatigue is alleviated by rest, while sleepiness is relieved by sleep […]. Unfortunately, however, fatigue and sleepiness may sometimes be confounded in clinical practice, research, and psychometry" (p. 38).

Lastly, in addition to investigating what symptoms are associated with depression and related mental disorders, we should aim to measure this information with higher precision. While current approaches to asking participants in binary or ordinal questions how well they slept in the last two weeks may be quick and practical, a thorough assessment of actual hours of sleep, sleeping patterns, as well as behavioral and pathophysiological data will likely yield important complementary information.

In sum, a new empirical baseline for depression research is needed. To obtain this baseline, psychiatry should rephrase its question from "what symptoms indicate the underlying disorder depression" to "how central are symptoms in psychopathological symptom networks". In order to answer this question, it is necessary to start increasing quantity and quality of symptom assessment.

9.2 Network Clusters and Subtypes

Certain symptom configurations will be much more prevalent than others, and it is possible that they share etiological processes; for example, insomnia and fatigue may commonly co-occur, whereas insomnia and hypersomnia may not. There is undoubtedly a need to classify and simplify, and the goal is not to establish every single symptom of all existing DSM conditions as distinct entity; we need useful categories and diagnoses; a symptom-based approach may ultimately provide symptom clusters and endophenotypes that are associated with specific gene polymorphisms or environmental predictor variables, and inform psychiatry about efficacious prevention and treatment strategies.

To do so, there are two alternative approaches. First, theory-driven perspectives to depression networks can be utilized, one of which is an evolutionary understanding of mood and low mood. As discussed in detail in three publications by Matthew Keller and colleagues (Keller et al., 2007; Keller & Nesse, 2005, 2006), particular life events may causally lead to specific symptom profiles of depression, because different forms of low mood may represent particular adaptive strategies to recurrent fitness threats in evolutionary environments (e.g., losing a partner or failing to reach an important goal). These theory-driven subtypes could be fit onto longitudinal network data (e.g., by constraining the effects of certain life events onto specific symptoms) to test their validity.

The second approach is data-driven and requires the development of exploratory clustering algorithms for symptom networks in order to detect meaningful subgroups. In a longitudinal epidemiological dataset, several causal symptom pathways into depression could be discovered, and secondary analyses could then reveal that some of these classes do share, for instance, genetic predispositions. While network clusters may not exist as real entities, they will likely prove to be much more useful than the current super-category MDD or its DSM-subtypes.

9.3 External Variables as Network Components

If symptoms are not indicators of a latent disorder, but constitute the disorder itself, the perspective on external variables such as demographic factors or life stress changes substantially: a specific life event such as the death of a loved one may be as central or even more central than depression symptoms in a network of relevant depression-related variables, extending the atheoretical DSM focus from symptoms to external factors. Including such variables also is sensible from the vantage point of HPCs: there is no intrinsic property that differentiates a depressive symptom from a life event in its function as a property of a given family like MDD; while symptoms are influential vari-

ables with significant causal power, some external variables may be equally important in the development and preservation of depression.

Self-esteem, for example, is not a depression symptom, but has been shown to be a very important factor in the development of depression. Self-esteem mediates the associations of child abuse and depression (Stein, Leslie, & Nyamathi, 2002), childhood trauma and depression (Turner & Butler, 2003), social support and depression (Symister & Friend, 2003), environmental risk for depression and depression (Prelow, Weaver, & Swenson, 2006), as well as parental conflict and depression (Turner & Kopiec, 2006). Furthermore, self-esteem interacts with failure to predict subsequent depressive reactions (Metalsky, Joiner, Hardin, & Abramson, 1993), and moderates the effects of attributional style on depression in times of stress (Metalsky et al., 1993; Robinson, Garber, & Hilsman, 1995)[35]. This example illuminates that a new models to psychopathological networks should shift the sole focus on symptoms and include external variables.

9.4 Longitudinal Extension of Network Approaches

Prior cross-sectional network research has established associations between symptoms (e.g., Cramer et al., 2011, 2010), but the causality of symptom effects remains unclear; to my knowledge, only one longitudinal analysis of symptom networks has been published so far (Bringmann et al., 2013). In this report, the authors utilized time-series data of 129 individuals across 12 days (6 days baseline, 6 days post-intervention) with a maximum of 60 measurements per day. All participants reported depressive symptoms at baseline and were assigned to either a treatment group or a control group, and roughly 14 weeks passed between the assessment of the first and the second 6 days. Two well-established statistical approaches were combined in the report. For the within-person analyses, a multivariate extension of the auto-regressive model was employed (vector autoregressive modeling; VAR), leading to an averaged population network of items such as 'cheerful', 'worry', 'fearful', 'sad', and 'relaxed'. This network represents how symptoms are associated with each other in the whole sample, and revealed that, for instance, 'fearful' and 'sad' affected each other across time. In addition, a multi-level approach was used to capture individual differences of symptom networks, leading to one symptom network per person. These personalized networks were then combined to construct a model in which the degree of individual variability for each connection in the network could be estimated. For example, the authors identified large inter-individual variability for the causal power of 'fear' on 'worry'.

[35] Most of the research cited in this paragraph was conducted with children and adolescents.

While the study only covered a few days, time-series data offer the opportunity to understand the development of depression networks across time and allow for investigating the impact of certain risk factors at certain time points. It is entirely possible that a child or adolescent may be more vulnerable to particular risk factors in specific developmental stages, potentially informing us about important prevention and intervention strategies. The publication of Bringmann et al. (2013) contains the syntax of their statistical approach to analyzing longitudinal networks, and will hopefully find broad application in the future.

9.5 Accessibility of Symptom-Based Data

Recently, there has been a trend to make datasets available for other researchers and the public. The STAR*D data used for Study 3 (Chapter ›7), for instance, was obtained from the NIMH. However, the majority of datasets are not available, and those that are often do not contain information on individual symptoms.

Personally, I have tried to obtain datasets of clinical trials in the last years testing the efficacy of antidepressant drugs. The largest repository of clinical trials, however – the database of the FDA – only contains aggregate data on sum-scores of depressive symptoms. As lined out in detail in Chapter ›4.4, specific antidepressants likely have both positive and negative impacts on particular MDD symptoms, and possibly highly relevant information is obfuscated by current practices. Hopefully, policy makers can soon be convinced that symptom-based approaches promise substantial benefits and will pave the way for making more symptom-level data available.

9.6 Nomenclature

When trying to collect research on individual depressive symptoms, mainly for the overview (Chapter ›4) as well as the studies in Chapters ›5, ›6, and ›7, it was surprisingly difficult to gather the current state of knowledge. This is primarily due to two reasons. First, the large majority of studies with the term "depressive symptoms" in either title or abstract do not actually contain analyses of individual symptoms; instead, sum-scores are presented. Second, most studies that do report symptom-based findings only describe them in secondary results, for example, in a few sentences in the results or discussion sections. Consequently, symptom-based reports are very difficult to find in keyword searches.

This means that psychiatry is not only faced with the problem of covert heterogeneity – it is also difficult to *find* studies on covert heterogeneity due to problems in nomenclature, since journals use the keyword "depressive symptoms" for studies on sum-scores, and authors often use

both terms interchangeably. It would be important to find a new and distinct tag that makes actual studies on individual psychopathological symptoms easier to identify within the vast amount of depression literature on sum-scores.

9.7 Sum-Scores

Un-weighted sum-scores that are the basis of any kind of diagnosis rely on the assumption of equivalent symptoms and one underlying latent variable. When thinking of symptoms as causally autonomous and closely interconnected variables, however, symptom centrality varies, and while some symptoms may be highly interwoven and have strong ties to a plethora of other symptoms, others may be fairly isolated (see Figure 8-1). The very idea behind clinical screening instruments of depression is to assess a latent disorder, and without such a latent condition – or with a condition that is extremely heterogeneous and fuzzy – sum-scores are not very useful.

At present, a cure for these problems associated with symptom sum-scores is not in sight, and the analysis of more meaningful units (individual symptoms) should be encouraged. Future research will be necessary to find alternative ways to build both practical and meaningful representations of overall symptom severity without oversimplifying pronounced covert heterogeneity.

9.8 Implications for Other Mental Disorders

Lastly, the question remains how well problems described in this dissertation generalize to other psychiatric disorders.

9.8.1 The pervasive problem of valid categories in psychiatry

It is widely acknowledged that diagnostic categories for mental disorders are problematic for numerous reasons, and many difficulties pertaining to depression may generalize to other psychiatric diseases; above all, diagnoses of mental disorders generally suffer from problems of validity.

In April 2013 – shortly before the release of the DSM-5 – Thomas Insel, director of the NIMH, published a statement in which he declared that "NIMH will be re-orienting its research away from DSM categories" (Insel, 2013); Insel further wrote that the reason DSM diagnoses will no longer be accepted as gold standard of psychiatric research is due to their lack of validity. Nature editor David Adam recently described the same problem in his statement that "[…] biologists have been unable to find any genetic or neuroscientific evidence to support the breakdown of complex mental disorders into separate categories" (Adam, 2013, p. 416), and the introduction of a leading psychiatric textbook reads: "there is little reason to believe that these categories are valid" (Grebb & Carlsen, 2009).

This pervasive problem of validity is reflected in various domains such as high comorbidity rates as well as lack of treatment specificity, associated biomarkers, and diagnostic stability. The editors of the "Research Agenda to the DSM-5" (Kupfer, First, & Regier, 2002) summarized the problems as follows (p. XVIII):

> "In more than 30 years since the introduction of the Feighner criteria […], the goal of validating these syndromes and discovering common etiologies has remained elusive […]. Despite many proposed candidates, not one laboratory marker has been found to be specific in identifying any of the DSM-defined syndromes. Epidemiological and clinical studies have shown extremely high rates of comorbidities among the disorders, undermining the hypothesis that the syndromes represent distinct etiologies. Furthermore, epidemiological studies have shown a high degree of short-term diagnostic instability for many disorders. With regard to treatment, lack of treatment specificity is the rule rather than the exception."

Similar to depression, most other DSM diagnoses are likely not distinct and clearly demarcated states, and many difficulties such as covert heterogeneity, problematic ideas of causation between a latent disorder and its indicators, and the assumption of local independence of symptoms may be important topics and deserve closer examination. While I can only speculate, it is possible that the notion of one common cause and the use of reflective models may fit other diagnoses such as PTSD better than depression; however, it is plausible that symptoms nonetheless directly interact with each other beyond a common cause even for such disorders. In sum, investigating symptom networks promises important insights for the majority of mental health conditions listed in the DSM-5.

9.8.2 Usefulness of psychiatric diagnoses

On the other hand, it is important to acknowledge the importance of psychiatric categories – they are necessary for standardizing research and treatment. From this point of view we have to investigate the *usefulness* of current DSM categories, and I suggest that psychiatric diagnoses may differ from each other in aspects that moderate their usefulness.

Especially Peter Zachar has championed a pragmatic approach to categorizing mental disorders, and he argues that "there are many different things in the DSM-IV and the ICD-10, and they cannot neatly be classified as the same type of thing" (Zachar, 2002, p. 219). This instrumentalist perspective on diagnoses describes *pragmatic kinds* – in contrast to natural or socially con-

structed kinds – and proponents of this pragmatic approach would argue that the property "real" does not add anything substantial to a psychiatric category (Fine, 1984)[36].

If one is confronted with a repository of several thousand scientific books, there is no correct system to bring order into the chaos, no real underlying structure beneath the multi-dimensional data-space. However, there is likely a way to organize the books that is *most useful* for the purpose at hand, e.g., for finding a subset of literature quickly that gives an overview pertaining to a specific topic. Alternatively, the books could be exhibited in a museum of modern art, in which case sorting them by size or color may be the most useful categorization.

As a concluding thought, I hypothesize that two important variables may moderate the usefulness of diagnostic categories: the degrees of etiological and symptomatological heterogeneity. A category is useful in this sense if it allows for a more reliable diagnosis and subsequently leads to more efficacious treatment of psychiatric conditions. Figure 9-1 shows a two-dimensional plane defined by these two dimensions.

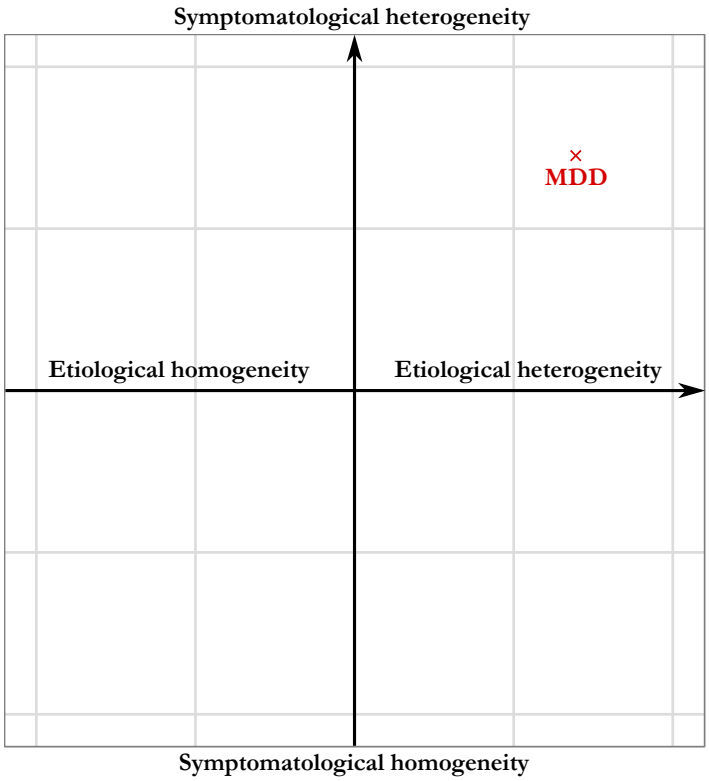

Figure 9-1. *Usefulness of diagnostic categories.*

[36] While pragmatic kinds are sometimes seen as alternative to HPCs (Kendler et al., 2010), the two theories can also be understood to complement each other: disorders are HPCs, and a pragmatic approach can be utilized for purposes of categorization and classification.

The four areas represent disorders that are symptomatically and etiologically heterogeneous (top right), symptomatically heterogeneous but etiologically homogenous (top left), symptomatically homogenous but etiologically heterogeneous (bottom right), and symptomatically and etiologically homogenous (bottom left). The most useful diagnosis is likely one of pronounced homogeneity both in terms of etiology and symptomatology. Not only does this allow for a more reliable diagnosis, which would enable researchers and clinicians to develop and implement more specific and well-directed treatment strategies, and particular etiologies would increase the probability of early detection and prevention. Moreover, the approach would increase the likelihood to discover distinct pathophysiological processes underlying mental disorders.

Depression may well be one of the most heterogeneous disorders on both dimensions; this makes MDD a great exemplary disorder for research on covert heterogeneity, and justifies further efforts of establishing more homogenous subtypes in order to improve treatment efficacy.

9.9 Conclusions

After many decades of depression research, core problems are unresolved. To unravel crucial questions pertaining to disease nature, disease etiology, and disease comorbidity, it is time to adopt a new framework that acknowledges the highly heterogeneous nature of MDD. Psychiatry should assess more symptoms – pathological and nonpathological symptoms of both medical and psychiatric conditions alike – and investigate the lawful connections of these symptoms and external variables such as life stress, personality traits, or genetic dispositions.

Scientific realism has recently led to the development of new symptom-based approaches. These models are highly multidimensional and complicated, and their application will take substantial amounts of research and time. But as Tukey pointed out:

> "Clarity in the large comes from clarity in the medium scale; clarity in the medium scale comes from clarity in the small. Clarity always comes with difficulty." (Tukey, 1969, p. 88)

I firmly believe that clarity derived from new network models will empower us to develop more efficient prevention and intervention strategies for one of the greatest mental health challenges of our time: Major Depressive Disorder.

10 References

Abramson, L. Y., Metalsky, G. I., & Alloy, L. B. (1989). Hopelessness depression: A theory-based subtype of depression. *Psychological Review*, *96*(2), 358–372. doi:10.1037//0033-295X.96.2.358

Achermann, P. (2004). The two-process model of sleep regulation revisited. *Aviation, space, and environmental medicine*, *75*(3 Suppl), A37–43.

Adam, D. (2013). On the spectrum. *Nature*, *496*(7446), 416–418. doi:10.1038/496416a

Aggen, S. H., Neale, M. C., & Kendler, K. S. (2005). DSM criteria for major depression: evaluating symptom patterns using latent-trait item response models. *Psychological Medicine*, *35*(4), 475–487. doi:10.1017/S0033291704003563

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. doi:10.1109/TAC.1974.1100705

Akiskal, H. S., & McKinney, W. T. (1973). Depressive Disorders: Toward a Unified Hypothesis. *Science*, *182*(4107), 20–29.

American Psychiatric Association. (1980). *Diagnostic and Statistical Manual of Mental Disorders, Third Edition*. Washington, DC: American Psychiatric Association.

American Psychiatric Association. (1994). *The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013a). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013b). Psychiatric News Alert. Retrieved from alert.psychiatricnews.org

Anderson, I. M., & Tomenson, B. M. (1995). Treatment discontinuation with selective serotonin reuptake inhibitors compared with tricyclic antidepressants: a meta-analysis. *BMJ*, *310*(6992), 1433–8.

Andrews, G., Slade, T., Sunderland, M., & Anderson, T. (2007). Issues for DSM-V: simplifying DSM-IV to enhance utility: the case of major depressive disorder. *The American journal of psychiatry*, *164*(12), 1784–5. doi:10.1176/appi.ajp.2007.07060928

Angst, J. (1999). Major depression in 1998: are we providing optimal therapy? *The Journal of clinical psychiatry*, *60 Suppl 6*, 5–9.

Angst, J., & Clayton, P. (1986). Premorbid personality of depressive, bipolar, and schizophrenic patients with special reference to suicidal issues. *Comprehensive psychiatry*, *27*(6), 511–32.

Baas, K. D., Cramer, A. O. J., Koeter, M. W. J., van de Lisdonk, E. H., van Weert, H. C., & Schene, A. H. (2011). Measurement invariance with respect to ethnicity of the Patient Health Questionnaire-9 (PHQ-9). *Journal of affective disorders*, *129*(1-3), 229–35. doi:10.1016/j.jad.2010.08.026

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2004). Reviews and Overviews The Hamilton Depression Rating Scale: Has the Gold Standard Become a Lead Weight ? *American Journal of Psyc*, *161*(12), 2163–2177.

Baglioni, C., Battagliese, G., Feige, B., Spiegelhalder, K., Nissen, C., Voderholzer, U., … Riemann, D. (2011). Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. *Journal of affective disorders*, *135*(1-3), 10–9. doi:10.1016/j.jad.2011.01.011

Baldwin, D. (2003). Essential considerations when choosing a modern antidepressant. *International Journal of Psychiatry in Clinical Practice*, *7*(s1), 3–8. doi:10.1080/13651500310000825

Bannister, D. (1968). The logical requirements of research into schizophrenia. *The British Journal of Psychiatry*, *114*(507), 181–8.

Barlow, D. H., Allen, L. B., & Choate, M. L. (2004). Toward a unified treatment for emotional disorders. *Behavior Therapy*, *35*(2), 205–230. doi:10.1016/S0005-7894(04)80036-4

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes.

Baumeister, H., & Parker, J. D. (2012). Meta-review of depressive subtyping models. *Journal of affective disorders*, *139*(2), 126–40. doi:10.1016/j.jad.2011.07.015

Bech, P. (2010). Struggle for subtypes in primary and secondary depression and their mode-specific treatment or healing. *Psychotherapy and psychosomatics*, *79*(6), 331–8. doi:10.1159/000320118

Beck, A., Brown, G., & Berchick, R. (1990). Relationship between hopelessness and ultimate suicide: a replication with psychiatric outpatients. *American Journal of Psychiatry*, *147*(2), 190–195.

Beck, A., Steer, R., & Garbin, M. (1988). Psychometric properties of the Beck Depression Inventory: 25 years of evaluation. *Clinical psychology review*, *8*, 77–100.

Beck, A. T., Brown, G., & Steer, R. A. (1989). Prediction of eventual suicide in psychiatric inpatients by clinical ratings of hopelessness. *Journal of consulting and clinical psychology*, *57*(2), 309–10.

Beck, A. T., Rush, J. A., Shaw, F. S., & Emery, G. (1979). *Cognitive Therapy of Depression*. New York: Guilford Press.

Beekman, A. T. F., Deeg, D. J., van Tilburg, T., Smit, J. H., Hooijer, C., & van Tilburg, W. (1995). Major and minor depression in later life: a study of prevalence and risk factors. *Journal of affective disorders*, *36*(1-2), 65–75.

Benca, R. M., & Peterson, M. J. (2008). Insomnia and depression. *Sleep medicine*, *9 Suppl 1*, S3–9. doi:10.1016/S1389-9457(08)70010-8

Berman, A. L. (2009). Depression and suicide. In Ian H Gotlib & C. L. Hammen (Eds.), *Handbook of depression* (2nd ed., pp. 510–530). New York, NY: Guilford Press.

Bet, P. M., Hugtenburg, J. G., Penninx, B. W. J. H., & Hoogendijk, W. J. G. (2013). Side effects of antidepressants during long-term use in a naturalistic setting. *European neuropsychopharmacology*, (2013), 1–9. doi:10.1016/j.euroneuro.2013.05.001

Blazer, D. G., Landerman, L. R., Hays, J. C., Simonsick, E. M., & Saunders, W. B. (1998). Symptoms of depression among community-dwelling elderly African-American and white older adults. *Psychological medicine*, *28*(6), 1311–20.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*(4-5), 175–308. doi:10.1016/j.physrep.2005.10.009

Bollen, K. A. (1989). *Structural Equations with Latent Variables* (p. 514). New York: Wiley.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological bulletin*, *110*(2), 305–314.

Bollini, P., Pampallona, S., Tibaldi, G., Kupelnick, B., & Munizza, C. (1999). Effectiveness of antidepressants. Meta-analysis of dose-effect relationships in randomised clinical trials. *The British Journal of Psychiatry*, *174*, 297–303.

Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature reviews. Genetics*, *3*(11), 872–82. doi:10.1038/nrg932

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of clinical psychology*, *64*(9), 1089–1108. doi:10.1002/jclp

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology*, *9*, 91–121. doi:10.1146/annurev-clinpsy-050212-185608

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PloS one*, *6*(11), e27407. doi:10.1371/journal.pone.0027407

Bothwell, S., & Weissman, M. M. (1977). Social impairments four years after an acute depressive episode. *The American journal of orthopsychiatry*, *47*(2), 231–7.

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), *Species: New Interdisciplinary Essays* (pp. 141–185). Cambridge, MA: A Bradford Book/The MIT Press.

Boyd, Richard. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical studies*, 127–148.

Breslau, N., Roth, T., Rosenthal, L., & Andreski, P. (1996). Sleep disturbance and psychiatric disorders: a longitudinal epidemiological study of young adults. *Biological psychiatry*, *39*(6), 411–8.

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., … Tuerlinckx, F. (2013). A network approach to psychopathology: new insights into clinical longitudinal data. *PloS one*, *8*(4), e60188. doi:10.1371/journal.pone.0060188

Brooks, D. (2013). Heroes of Uncertainty. *The New York Times*. Retrieved from http://www.nytimes.com/2013/05/28/opinion/brooks-heroes-of-uncertainty.html?_r=2&

Brown, C., Schulberg, H. C., & Madonia, M. J. (1995). Assessing Depression in Primary Care Practice With the Beck Depression Inventory and the Hamilton Rating Scale for Depression. *Psychological Assessment*, *7*(1), 59–65.

Brown, G. K., Beck, A. T., Steer, R. A., & Grisham, J. R. (2000). Risk factors for suicide in psychiatric outpatients: a 20-year prospective study. *Journal of consulting and clinical psychology*, *68*(3), 371–7.

Brown, G. W., & Harris, T. O. (1989). Depression. In George W Brown & T. O. Harris (Eds.), *Life Events and Illness* (pp. 49–93). New York, NY: Guilford Press.

Brown, G W, Harris, T. O., & Hepworth, C. (1994). Life Events and Endogenous Depression. *Archives of general psychiatry*, *51*, 525–534.

Brown, G W, Nibhrolchain, M., & Harris, T. (1979). Psychotic and neurotic depression: Part 3. Aetiological and background factors. *Journal of Affective Disorders*, *1*, 195–211.

Brown, W. A. (2007). Treatment response in melancholia. *Acta psychiatrica Scandinavica. Supplementum*, (433), 125–9. doi:10.1111/j.1600-0447.2007.00970.x

Butterfield, P. S. (1988). The stress of residency. A review of the literature. *Archives of internal medicine*, *148*(6), 1428–35.

Cannon, T. D., & Keller, M. C. (2006). Endophenotypes in the genetic analyses of mental disorders. *Annual review of clinical psychology*, *2*, 267–90. doi:10.1146/annurev.clinpsy.2.022305.095232

Cassidy, W. L., Planagan, N. B., Spellman, M., & Cohen, M. E. (1957). Clinical observations in manic-depressive disease; a quantitative study of one hundred manic-depressive patients and fifty medically sick controls. *Journal of the American Medical Association*, *164*(14), 1535–46.

Cho, H. J., Lavretsky, H., Olmstead, R., Levin, M. J., Oxman, M. N., & Irwin, M. R. (2008). Sleep disturbance and depression recurrence in community-dwelling older adults: a prospective study. *The American journal of psychiatry*, *165*(12), 1543–50. doi:10.1176/appi.ajp.2008.07121882

Christensen, R. (2012). ordinal: regression models for ordinal data. Retrieved from http://cran.r-project.org/web/packages/ordinal

Cohen, L. H., Towbes, L. C., & Flocco, R. (1988). Effects of induced mood on self-reported life events and perceived and received social support. *Journal of personality and social psychology*, *55*(4), 669–74.

Colman, I., Naicker, K., Zeng, Y., Ataullahjan, A., Senthilselvan, A., & Patten, S. B. (2011). Predictors of long-term prognosis of depression. *Canadian Medical Association journal*, *183*(17), 1955–6. doi:10.1503/cmaj.111664

Coryell, W., Winokur, G., Maser, J. D., Akiskal, H. S., Keller, M. B., & Endicott, J. (1994). Recurrently situational (reactive) depression: A study of course, phenomenology and familial psychopathology. *Journal of Affective Disorders*, *31*(3), 203–210. doi:10.1016/0165-0327(94)90030-2

Costa, P. T., & McCrae, R. R. (1997). Stability and change in personality assessment: the revised NEO Personality Inventory in the year 2000. *Journal of personality assessment*, *68*(1), 86–94. doi:10.1207/s15327752jpa6801_7

Costello, C. (1993). The advantages of the symptom approach to depression. In C. Costello (Ed.), *Symptoms of Depression* (pp. 1–21). New York: John Wiley and Sons.

Cramer, A. O. J., Borsboom, D., Aggen, S. H., & Kendler, K. S. (2011). The pathoplasticity of dysphoric episodes: differential impact of stressful life events on the pattern of depressive symptom inter-correlations. *Psychological medicine*, *42*(5), 957–65. doi:10.1017/S003329171100211X

Cramer, A. O. J., Kendler, K. S., & Borsboom, D. (2011). Discussion on "Heritability in the Era of Molecular Genetics: Some Thoughts for Understanding Genetic Influences on Behavioural Traits." *European Journal of Personality*, *286*(May), 270–271. doi:10.1002/per

Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: a network perspective. *The Behavioral and brain sciences*, *33*(2-3), 137–50; discussion 150–93. doi:10.1017/S0140525X09991567

Crockett, L. J., Randall, B. A., Shen, Y.-L., Russell, S. T., & Driscoll, A. K. (2005). Measurement equivalence of the center for epidemiological studies depression scale for Latino and Anglo adolescents: a national study. *Journal of consulting and clinical psychology*, *73*(1), 47–58. doi:10.1037/0022-006X.73.1.47

David, M., Ceschi, G., Billieux, J., & Van der Linden, M. (2008). Depressive symptoms after trauma: is self-esteem a mediating factor? *The Journal of nervous and mental disease*, *196*(10), 735–42. doi:10.1097/NMD.0b013e3181879dd8

Davidson, J. R. T. (2007). A history of the concept of atypical depression. *The Journal of clinical psychiatry*, *68*(Suppl 3), 10–15.

Dawson, D., & Reid, K. (1997). Fatigue, alcohol and performance impairment. *Nature*, *388*(6639), 235. doi:10.1038/40775

Dew, M. A., Reynolds, C. F., Houck, P. R., Hall, M., Buysse, D. J., Frank, E., & Kupfer, D. J. (1997). Temporal profiles of the course of depression during treatment. Predictors of pathways toward recovery in the elderly. *Archives of general psychiatry*, *54*(11), 1016–24.

Dombrovski, A. Y., Mulsant, B. H., Houck, P. R., Mazumdar, S., Lenze, E. J., Andreescu, C., … Reynolds, C. F. (2007). Residual symptoms and recurrence during maintenance treatment of late-life depression. *Journal of affective disorders*, *103*(1-3), 77–82. doi:10.1016/j.jad.2007.01.020

Duffy, T. P. (2005). Glory days. What price glory? *The Pharos of Alpha Omega Alpha-Honor Medical Society. Alpha Omega Alpha*, *68*(4), 22–30.

Durmer, J., & Dinges, D. (2005). Neurocognitive consequences of sleep deprivation. *Seminars in neurology*, *25*(1), 117–129.

Ekman, P., & Cordaro, D. (2011). What is Meant by Calling Emotions Basic. *Emotion Review*, *3*(4), 364–370. doi:10.1177/1754073911410740

Emmelkamp, P. M. G. (2003). Behavior therapy with adults. In M. Lambert (Ed.), *Handbook of psychotherapy and behavior change* (5dh editio., pp. 393–446). New York, NY: Wiley.

Endicott, J., Nee, J., Harrison, W., & Blumenthal, R. (1993). Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. *Psychopharmacology bulletin*, *29*(2), 321–6.

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, *48*(4), 1–18.

Ereshefsky, M. (2007). Species. In E. N. Zalte (Ed.), *The Stanford Encyclopedia of Philosophy (Summer 2007 Edition)*.

Fairclough, S. H., & Graham, R. (1999). Impairment of driving performance caused by sleep deprivation or alcohol: a comparative study. *Human factors*, *41*(1), 118–28.

Fava, M., Rush, A. J., Alpert, J. E., Balasubramani, G. K., Wisniewski, S. R., Carmin, C. N., … Trivedi, M. H. (2008). Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. *The American journal of psychiatry*, *165*(3), 342–51. doi:10.1176/appi.ajp.2007.06111868

Fava, M., Rush, A. J., Trivedi, M. H., Nierenberg, A., Thase, M. E., Sackeim, H. A., … Kupfer, D. J. (2003). Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *The Psychiatric clinics of North America*, *26*(2), 457–94.

Fawcett, J, Scheftner, W. A., Fogg, L., Clark, D. C., Young, M. A., Hedeker, D., & Gibbons, R. (1990). Time-related predictors of suicide in major affective disorder. *The American journal of psychiatry*, *147*(9), 1189–94.

Fawcett, Jan, Scheftner, W., & Clark, D. (1987). Clinical predictors of suicide in patients with major affective disorders: a controlled prospective study. *American Journal of Psychiatry*, *144*(1), 35–40.

Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Winokur, G., & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of general psychiatry*, *26*(1), 57–63.

Ferentinos, P., Kontaxakis, V., Havaki-Kontaxaki, B., Paparrigopoulos, T., Dikeos, D., Ktonas, P., & Soldatos, C. (2009). Sleep disturbances in relation to fatigue in major depression. *Journal of psychosomatic research*, *66*(1), 37–42. doi:10.1016/j.jpsychores.2008.07.009

Fine, A. (1984). The natural ontological attitude. In J. Leplin (Ed.), *Scientific Realism* (pp. 83–107). Berkeley, CA: University of California Press.

Ford, D. E., & Kamerow, D. B. (1989). Epidemiologic study of sleep disturbances and psychiatric disorders. An opportunity for prevention? *JAMA : the journal of the American Medical Association*, *262*(11), 1479–84.

Friedman, R. A. (2013). Grief, Depression, and the DSM-5. *The New England journal of medicine*, *366*(20), 1855–57.

Furukawa, T. A., McGuire, H., & Barbui, C. (2002). Meta-analysis of effects and side effects of low dosage tricyclic antidepressants in depression: systematic review. *BMJ*, *325*(7371), 991.

Furukawa, T., Streiner, D., Azuma, H., Higuchi, T., Kamijima, K., Kanba, S., … Miura, S. (2005). Cross-cultural equivalence in depression assessment: Japan-Europe-North American study. *Acta psychiatrica Scandinavica*, *112*(4), 279–85. doi:10.1111/j.1600-0447.2005.00587.x

Gelman, S. a. (2009). Learning from others: children's construction of concepts. *Annual review of psychology*, *60*, 115–40. doi:10.1146/annurev.psych.59.103006.093659

Gilmer, W. S., & McKinney, W. T. (2003). Early experience and depressive disorders: human and non-human primate studies. *Journal of affective disorders*, *75*(2), 97–113.

Goldberg, D. (2011). The heterogeneity of "major depression." *World Psychiatry*, *10*(3), 226–228.

Goldston, K., & Baillie, A. J. (2008). Depression and coronary heart disease: a review of the epidemiological evidence, explanatory mechanisms and management approaches. *Clinical psychology review*, *28*(2), 288–306. doi:10.1016/j.cpr.2007.05.005

Gotlib, I H, Lewinsohn, P. M., & Seeley, J. R. (1995). Symptoms versus a diagnosis of depression: differences in psychosocial functioning. *Journal of consulting and clinical psychology*, *63*(1), 90–100.

Gottesman, I. I., & Gould, T. D. (2003). Reviews and Overviews The Endophenotype Concept in Psychiatry: Etymology and Strategic Intentions. *American Journal of Psychiatry*, *160*(April), 636–645.

Grebb, J. A., & Carlsen, A. (2009). Introduction and considerations for a brain-based diagnostic system in psychiatry. In B. Sadock, V. Sadock, P. Ruiz, & H. Kaplan (Eds.), *Comprehensive Textbook of Psychiatry* (9th editio., pp. 1–4). Philadelphia: Wolters Kluwer Health/Lippincott Williams.

Greer, T. L., Kurian, B. T., & Trivedi, M. H. (2010). Defining and measuring functional recovery from depression. *CNS drugs*, *24*(4), 267–84. doi:10.2165/11530230-000000000-00000

Grömping, U. (2006). Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software*, *17*(1), 1–27.

Grömping, U. (2007). Estimators of Relative Importance in Linear Regression Based on Variance Decomposition. *The American Statistician*, *61*(2), 139–147. doi:10.1198/000313007X188252

Gros, D. F., Price, M., Magruder, K. M., & Frueh, B. C. (2012). Symptom overlap in posttraumatic stress disorder and major depression. *Psychiatry research*, *196*(2-3), 267–70. doi:10.1016/j.psychres.2011.10.022

Gutman, D. A., & Nemeroff, C. B. (2003). Persistent central nervous system effects of an adverse early environment: clinical and preclinical studies. *Physiology & behavior*, *79*(3), 471–8.

Haack, S. (2003). *Defending Science –Within Reason: Between Scientism and Cynicism*. Amherst, NY: Prometheus Books.

Hamilton, M. (1960). A rating scale for depression. *Journal of neurology, neurosurgery, and psychiatry*, *23*, 56–62.

Hammen, C. (2005). Stress and depression. *Annual review of clinical psychology*, *1*, 293–319. doi:10.1146/annurev.clinpsy.1.102803.143938

Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: a review. *Journal of experimental psychology. Applied*, *6*(3), 236–49.

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *The British journal of social psychology / the British Psychological Society*, *39 (Pt 1)*, 113–27.

Hasler, G, & Northoff, G. (2011). Discovering imaging endophenotypes for major depression. *Molecular psychiatry*, *16*(6), 604–19. doi:10.1038/mp.2011.23

Hasler, Gregor, Drevets, W. C., Manji, H. K., & Charney, D. S. (2004). Discovering endophenotypes for major depression. *Neuropsychopharmacology*, *29*(10), 1765–81. doi:10.1038/sj.npp.1300506

Hayduk, L. A., Robinson, H. P., & Perks, T. (2007). The weird world, and equally weird measurement models: reactive indicators and the validity revolution. *Structural Equation Modeling*, *14*(2), 280–310.

Hays, R. D., Wells, K. B., Sherbourne, C. D., Rogers, W., & Spritzer, K. (1995). Functioning and well-being outcomes of patients with depression compared with chronic general medical illnesses. *Archives of general psychiatry*, *52*(1), 11–9. doi:10.1001/archpsyc.1995.03950130011002

Heiberger, R. M., & Holland, B. (2004). *Statistical Analysis and Data Display: An Intermediate Course with Examples in S-PLUS, R, and SAS*. Springer.

Hek, K., Demirkan, A., Lahti, J., & Terracciano, A. (2013). A Genome-Wide Association Study of Depressive Symptoms. *Biological psychiatry*, *73(7)*, 667–78. doi:10.1016/j.biopsych.2012.09.033

Helmes, E., & Nielson, W. R. (1998). An examination Epidemiological of the internal structure of the Center for Studies-Depression Scale in two medical samples. *Personality and Individual Differences*, *25*, 735–743.

Henry, B., Moffitt, T. E., Caspi, A., Langley, J., & Silva, P. A. (1994). On the "remembrance of things past": A longitudinal evaluation of the retrospective method. *Psychological Assessment*, *6*(2), 92–101.

Hirschfeld, R. M. (1981). Situational depression: validity of the concept. *The British Journal of Psychiatry*, *139*(4), 297–305. doi:10.1192/bjp.139.4.297

Hirschfeld, R. M., Dunner, D. L., Keitner, G., Klein, D. N., Koran, L. M., Kornstein, S. G., … Keller, M. B. (2002). Does psychosocial functioning improve independent of depressive symptoms? A comparison of nefazodone, psychotherapy, and their combination. *Biological psychiatry*, *51*(2), 123–33.

Hirschfeld, R. M., Montgomery, S. A., Keller, M. B., Kasper, S., Schatzberg, A. F., Möller, H. J., … Bourgeois, M. (2000). Social functioning in depression: a review. *The Journal of clinical psychiatry*, *61*(4), 268–75.

Holland, P., & Rosenbaum, P. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics*, *14*(4), 1523–1543.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. doi:10.1080/10705519909540118

Hyland, M. E. (2011). *The Origins of Health and Disease*. Cambridge, UK: Cambridge University Press.

IBM Corp. (2012). IBM SPSS Statistics, Version 21.0. Armonk, NY: IBM Corp.

Insel, T. R. (2013). Transforming Diagnosis. *National Institute of Mental Health*. Retrieved from http://www.nimh.nih.gov/about/director/2013/transforming-diagnosis.shtml

Iwata, N., & Roberts, R. E. (1996). Age differences among Japanese on the Center for Epidemiologic Studies Depression Scale: an ethnocultural perspective on somatization. *Social science & medicine (1982)*, *43*(6), 967–74.

James, W. (1890). *The principles of psychology [re-published in 1950]*. New York, NY: Dover Publications.

Jang, K. L., Livesley, W. J., Taylor, S., Stein, M. B., & Moon, E. C. (2004). Heritability of individual depressive symptoms. *Journal of affective disorders*, *80*(2-3), 125–33. doi:10.1016/S0165-0327(03)00108-3

Johnson, E. O., Roth, T., & Breslau, N. (2006). The association of insomnia with anxiety disorders and depression: exploration of the direction of risk. *Journal of psychiatric research*, *40*(8), 700–8. doi:10.1016/j.jpsychires.2006.07.008

Johnson, J. W., & Lebreton, J. M. (2004). History and Use of Relative Importance Indices in Organizational Research. *Organizational Research Methods*, *7*(3), 238–257. doi:10.1177/1094428104266510

Johnson, W., Penke, L., & Spinath, F. (2011). Heritability in the era of molecular genetics: Some thoughts for understanding genetic influences on behavioural traits. *European Journal of …*, *266*(December 2010), 254–266. doi:10.1002/per

Jones, R. N. (2006). Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination. Detecting differential item functioning using MIMIC modeling. *Medical care*, *44*(11 Suppl 3), S124–33. doi:10.1097/01.mlr.0000245250.50114.0f

Jöreskog, K., & Goldberger, A. (1975). Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable. *Journal of the American Statistical Association*, *70*(351), 631–639.

Joyce, P. R., Mulder, R. T., McKenzie, J. M., Luty, S. E., & Cloninger, C. R. (2004). Atypical depression, atypical temperament and a differential antidepressant response to fluoxetine and nortriptyline. *Depression and anxiety*, *19*(3), 180–6. doi:10.1002/da.20001

Judd, L. L., Paulus, M. P., Wells, K. B., & Rapaport, M. H. (1996). Socioeconomic burden of subsyndromal depressive symptoms and major depression in a sample of the general population. *The American journal of psychiatry*, *153*(11), 1411–7.

Judd, L. L., Schettler, P. J., Coryell, W., Akiskal, H. S., & Fiedorowicz, J. G. (2013). Overt Irritability/Anger in Unipolar Major Depressive Episodes: Past and Current Characteristics and Implications for Long-term Course. *JAMA psychiatry*, *92093*. doi:10.1001/jamapsychiatry.2013.1957

Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular psychiatry*, *17*(12), 1174–1179. doi:10.1038/mp.2012.105

Katon, W. J. (2003). Clinical and health services relationships between major depression, depressive symptoms, and general medical illness. *Biological psychiatry*, *54*(3), 216–26.

Katschnig, H., Pakesch, G., & Egger-Zeidner, E. (1986). Life stress and depressive subtypes: A review of present diagnostic criteria and recent research results. In H. Katschnig (Ed.), *Life events and psychiatric disorders: Controversial issues*. Cambridge MA: Cambridge University Press.

Kaymaz, N., van Os, J., Loonen, A. J. M., & Nolen, W. A. (2008). Evidence that patients with single versus recurrent depressive episodes are differentially sensitive to treatment discontinuation: a meta-analysis of placebo-controlled randomized trials. *The Journal of clinical psychiatry*, *69*(9), 1423–36.

Keane, T. M., Brief, D. J., Pratt, E. M., & Miller, M. W. (2007). Assessment of PTSD and its comorbidities in adults. In M. J. Friedman, T. M. Keane, & P. A. Resick (Eds.), *Handbook of PTSD: Science and practice* (pp. 279–305). New York, NY: Guilford Press.

Keller, M. C., Neale, M. C., & Kendler, K. S. (2007). Association of different adverse life events with distinct patterns of depressive symptoms. *The American journal of psychiatry*, *164*(10), 1521–9; quiz 1622. doi:10.1176/appi.ajp.2007.06091564

Keller, M. C., & Nesse, R. M. (2005). Is low mood an adaptation? Evidence for subtypes with symptoms that match precipitants. *Journal of affective disorders*, *86*(1), 27–35. doi:10.1016/j.jad.2004.12.005

Keller, M. C., & Nesse, R. M. (2006). The evolutionary significance of depressive symptoms: different adverse situations lead to different depressive symptom patterns. *Journal of personality and social psychology*, *91*(2), 316–30. doi:10.1037/0022-3514.91.2.316

Kendell, R., & Jablensky, A. (2003). Distinguishing Between the Validity and Utility of Psychiatric Diagnoses. *American Journal of Psychiatry*.

Kendler, K S, & Gardner, C. O. (1998). Boundaries of major depression: an evaluation of DSM-IV criteria. *The American journal of psychiatry*, *155*(2), 172–7.

Kendler, K S, Zachar, P., & Craver, C. (2010). What kinds of things are psychiatric disorders? *Psychological medicine*, (September 2010), 1–8. doi:10.1017/S0033291710001844

Kendler, K., Gatz, M., Gardner, C. O., & Pedersen, N. L. (2006). A Swedish national twin study of lifetime major depression. *American Journal of Psychiatry*, *163*, 109–114.

Kendler, Kenneth S. (2005). "A gene for...": the nature of gene action in psychiatric disorders. *The American journal of psychiatry*, *162*(7), 1243–52. doi:10.1176/appi.ajp.162.7.1243

Kendler, Kenneth S, Aggen, S. H., & Neale, M. C. (2013). Evidence for Multiple Genetic Factors Underlying DSM-IV Criteria for Major Depression. *American Journal of Psychiatry*, *70(6)*, 599–607. doi:10.1001/jamapsychiatry.2013.751

Kendler, Kenneth S, & Baker, J. H. (2007). Genetic influences on measures of the environment: a systematic review. *Psychological medicine*, *37*(5), 615–26. doi:10.1017/S0033291706009524

Kendler, Kenneth S, Karkowski, L. M., & Prescott, C. A. (1999). Causal Relationship Between Stressful Life Events and the Onset of Major Depression. *American Journal of Psychiatry*, *156*, 837–841.

Kendler, Kenneth S, Kuhn, J., & Prescott, C. A. (2004). The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. *The American journal of psychiatry*, *161*(4), 631–6.

Kendler, Kenneth S, Myers, J., & Prescott, C. A. (2005). Sex differences in the relationship between social support and risk for major depression: a longitudinal study of opposite-sex twin pairs. *The American journal of psychiatry*, *162*(2), 250–6. doi:10.1176/appi.ajp.162.2.250

Kendler, Kenneth S, & Zachar, P. (2008). The incredible insecurity of psychiatric nosology. In Kenneth S Kendler & J. Parnas (Eds.), *Philosophical Issues in Psychiatry* (pp. 368–383). Baltimore, MD: The Johns Hopkins University Press.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., … Wang, P. S. (2003). The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA : the journal of the American Medical Association*, *289*(23), 3095–105. doi:10.1001/jama.289.23.3095

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of general psychiatry*, *62*(6), 617–27. doi:10.1001/archpsyc.62.6.617

Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General*, *131*(4), 451–476. doi:10.1037//0096-3445.131.4.451

Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS medicine*, *5*(2), e45. doi:10.1371/journal.pmed.0050045

Klonsky, E. D., Kotov, R., Bakst, S., Rabinowitz, J., & Bromet, E. J. (2012). Hopelessness as a predictor of attempted suicide among first admission patients with psychosis: a 10-year cohort study. *Suicide & life-threatening behavior*, *42*(1), 1–10. doi:10.1111/j.1943-278X.2011.00066.x

Knol, M. J., Twisk, J. W. R., Beekman, A. T. F., Heine, R. J., Snoek, F. J., & Pouwer, F. (2006). Depression as a risk factor for the onset of type 2 diabetes mellitus. A meta-analysis. *Diabetologia*, *49*(5), 837–45. doi:10.1007/s00125-006-0159-x

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, *16*(9), 606–13.

Kruskal, W., & Majors, R. (1989). Concepts of Relative Importance in Recent Scientific Literature. *The American Statistician*, *43*, 2–6.

Krystal, A. (2003). Insomnia in women. *Clinical cornerstone*, *5*(3), 41–48.

Kuo, W.-H., Gallo, J. J., & Eaton, W. W. (2004). Hopelessness, depression, substance disorder, and suicidality--a 13-year community-based study. *Social psychiatry and psychiatric epidemiology*, *39*(6), 497–501. doi:10.1007/s00127-004-0775-z

Kupfer, D., First, M., & Regier, D. (2002). *A research agenda for DSM V*. Washington, DC: American Psychiatric Association.

Laje, G., Allen, A. S., Akula, N., Manji, H., Rush, A. J., & Mcmahon, F. J. (2009). Genome-wide Association Study of Suicidal Ideation Emerging During Citalopram Treatment of Depressed Outpatients, *19*(9), 1–14. doi:10.1097/FPC.0b013e32832e4bcd.Genome-wide

Lam, R. W., & Stewart, J. N. (1996). The validity of atypical depression in DSM-IV. *Comprehensive psychiatry*, *37*(6), 375–83.

Laugsand, L. E., Strand, L. B., Platou, C., Vatten, L. J., & Janszky, I. (2013). Insomnia and the risk of incident heart failure: a population study. *European heart journal*, *Epub ahead*. doi:10.1093/eurheartj/eht019

Lee, S. W., Stewart, S. M., Byrne, B. M., Wong, J. P. S., Ho, S. Y., Lee, P. W. H., & Lam, T. H. (2008). Factor structure of the Center for Epidemiological Studies Depression Scale in Hong Kong adolescents. *Journal of personality assessment*, *90*(2), 175–84. doi:10.1080/00223890701845385

Lewinsohn, P. M., Klein, D. N., Durbin, E. C., Seeley, J. R., & Rohde, P. (2003). Family study of subthreshold depressive symptoms: risk factor for MDD? *Journal of affective disorders*, *77*(2), 149–57.

Lewis, C. M., Ng, M. Y., Butler, A. W., Cohen-Woods, S., Uher, R., Pirlo, K., … McGuffin, P. (2010). Genome-wide association study of major recurrent depression in the U.K. population. *The American journal of psychiatry*, *167*(8), 949–57. doi:10.1176/appi.ajp.2010.09091380

Lichstein, K., Wilson, N., & Johnson, C. (2000). Psychological treatment of secondary insomnia. *Psychology and aging*, *15*(2), 232–240. doi:10.1037//0882-7974.I5.2.232

Lichtenberg, P., & Belmaker, R. H. (2010). Subtyping major depressive disorder. *Psychotherapy and psychosomatics*, *79*(3), 131–5. doi:10.1159/000286957

Lieberman, J. (2013). Nothing to be defensive about. *Psychiatry News*. doi:10.1176/appi.pn.2013.7b19

Lindquist, K. a, Gendron, M., Oosterwijk, S., & Barrett, L. F. (2013). Do people essentialize emotions? Individual differences in emotion essentialism and emotional experience. *Emotion (Washington, D.C.)*, *13*(4), 629–44. doi:10.1037/a0032283

Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T., & Murray, C. J. L. (2006). Global Burden of Disease and Risk Factors. Washington (DC): World Bank.

Lux, V., & Kendler, K. S. (2010). Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria. *Psychological medicine*, *40*(10), 1679–90. doi:10.1017/S0033291709992157

Luyten, P., Blatt, S. J., Van Houdenhove, B., & Corveleyn, J. (2006). Depression research and treatment: are we skating to where the puck is going to be? *Clinical psychology review*, *26*(8), 985–99. doi:10.1016/j.cpr.2005.12.003

Ma, S. H., & Teasdale, J. D. (2004). Mindfulness-based cognitive therapy for depression: replication and exploration of differential relapse prevention effects. *Journal of consulting and clinical psychology*, *72*(1), 31–40. doi:10.1037/0022-006X.72.1.31

Marks, I. M. (1986). *Behavioural Psychotherapy* (Vol. 1986). Bristol: John Wright.

Martin, L. A., Neighbors, H. W., & Griffith, D. M. (2013). The Experience of Symptoms of Depression in Men vs Women: Analysis of the National Comorbidity Survey Replication. *JAMA psychiatry*, 1–7. doi:10.1001/jamapsychiatry.2013.1985

Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS medicine*, *3*(11), e442. doi:10.1371/journal.pmed.0030442

Mazure, C. M. (1998). Life Stressors as Risk Factors in Depression. *Clinical Psychology: Science and Practice*, *5*(3), 291–313. doi:10.1111/j.1468-2850.1998.tb00151.x

McClintock, S. M., Husain, M. M., Greer, T. L., & Cullum, C. M. (2010). Association between depression severity and neurocognitive function in major depressive disorder: a review and synthesis. *Neuropsychology*, *24*(1), 9–34. doi:10.1037/a0017336

Mcgill, B. C. (2011). *The biopsychosocial approach to understanding, subtyping, and treating depression: Results from the National Comorbidity Survey – replication*. University of North Texas.

McGlinchey, J. B., Zimmerman, M., Young, D., & Chelminski, I. (2006). Diagnosing major depressive disorder VIII: are some symptoms better than others? *The Journal of nervous and mental disease*, *194*(10), 785–90. doi:10.1097/01.nmd.0000240222.75201.aa

McGrath, P J, Stewart, J. W., Janal, M. N., Petkova, E., Quitkin, F. M., & Klein, D. F. (2000). A placebo-controlled study of fluoxetine versus imipramine in the acute treatment of atypical depression. *The American journal of psychiatry*, *157*(3), 344–50.

McGrath, Patrick J, Khan, A. Y., Trivedi, M. H., Stewart, J. W., Morris, D. W., Wisniewski, S. R., … Rush, A. J. (2008). Response to a selective serotonin reuptake inhibitor (citalopram) in major depressive disorder with melancholic features: a STAR*D report. *The Journal of clinical psychiatry*, *69*(12), 1847–55.

Mcgrath, R. E. (2005). Conceptual complexity and construct validity. *Journal of Personality Assessment*, *85*(2), 37–41. doi:10.1207/s15327752jpa8502

McKnight, P. E., & Kashdan, T. B. (2009). The importance of functional impairment to mental health outcomes: a case for reassessing our goals in depression treatment research. *Clinical psychology review*, *29*(3), 243–59. doi:10.1016/j.cpr.2009.01.005

Melartin, T., Leskelä, U., Rytsälä, H., Sokero, P., Lestelä-Mielonen, P., & Isometsä, E. (2004). Co-morbidity and stability of melancholic features in DSM-IV major depressive disorder. *Psychological Medicine*, *34*(08), 1443. doi:10.1017/S0033291704002806

Metalsky, G. I., Joiner, T. E., Hardin, T. S., & Abramson, L. Y. (1993). Depressive reactions to failure in a naturalistic setting: a test of the hopelessness and self-esteem theories of depression. *Journal of abnormal psychology*, *102*(1), 101–9.

Mineka, S., Watson, D., & Clark, L. A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual review of psychology*, *49*, 377–412. doi:10.1146/annurev.psych.49.1.377

Mitchell, A. J., McGlinchey, J. B., Young, D., Chelminski, I., & Zimmerman, M. (2009). Accuracy of specific symptoms in the diagnosis of major depressive disorder in psychiatric out-patients: data from the MIDAS project. *Psychological medicine*, *39*(7), 1107–16. doi:10.1017/S0033291708004674

Monroe, S. M., Harkness, K., Simons, A. D., & Thase, M. E. (2001). Life stress and the symptoms of major depression. *The Journal of nervous and mental disease*, *189*(3), 168–75.

Morin, A., & Ninot, G. (2011). Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in French clinical and nonclinical adults. *Revue d'Épidémiologie et de Santé Publique*, *59*, 327–340. doi:10.1016/j.respe.2011.03.061

Mossaheb, N., Zehetmayer, S., Jungwirth, S., Weissgram, S., Rainer, M., Tragl, K.-H., & Fischer, P. (2012). Are specific symptoms of depression predictive of Alzheimer's dementia? *The Journal of clinical psychiatry*, *73*(7), 1009–15. doi:10.4088/JCP.11m06962

Mouchabac, S., Ferreri, M., Cabanac, F., & Bitton, M. (2003). Residual symptoms after a treated major depressive disorder: a survey among private psychiatrists. *Encephale*, *29*(4), 306–12.

Mueller, T. I., Leon, A. C., Keller, M. B., Solomon, D. A., Endicott, J., Coryell, W., … Maser, J. D. (1999). Recurrence after recovery from major depressive disorder during 15 years of observational follow-up. *The American journal of psychiatry*, *156*(7), 1000–6.

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *The British Journal of Psychiatry*, *180*, 461–4. doi:10.1192/bjp.180.5.461

Murray, C. J. L., & Lopez, A. (1996). *Global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge MA: Harvard School of Public Health.

Murray, E. J. (1965). *Sleep, Dreams, and Arousal*. New York: Appleton-Century-Crofts.

Muscatell, K., Slavich, G., Monroe, S. M., & Gotlib, I. H. (2009). Stressful life events, chronic difficulties, and the symptoms of clinical depression. *The Journal of nervous and mental disease*, *197*(3), 154–160. doi:10.1097/NMD.0b013e318199f77b.Stressful

Muthén, B. O., & Muthén, L. (2012). *Mplus User's Guide, seventh edition*. Los Angeles: Muthén & Muthén.

Myung, W., Song, J., Lim, S.-W., Won, H.-H., Kim, S., Lee, Y., … Kim, D. K. (2012). Genetic association study of individual symptoms in depression. *Psychiatry research*, *198(3)*, 400–6. doi:10.1016/j.psychres.2011.12.037

National Institute of Health. (2003). *Breaking ground, breaking through: The Strategic Plan for Mood Disorders Research (NIH Publication No. 03–5121)*. Washington, DC: National Institute of Health.

Nesse, R. M., & Stein, D. J. (2012). Towards a genuinely medical model for psychiatric nosology. *BMC medicine*, *10*(1), 5. doi:10.1186/1741-7015-10-5

Nierenberg, A., Trivedi, M. H., Fava, M., Biggs, M. M., Shores-Wilson, K., Wisniewski, S. R., … Rush, A. J. (2007). Family history of mood disorder and characteristics of major depressive disorder: a STAR*D (sequenced treatment alternatives to relieve depression) study. *Journal of psychiatric research*, *41*(3-4), 214–21. doi:10.1016/j.jpsychires.2006.02.005

Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of abnormal psychology*, *109*(3), 504–11.

Okajima, I., Komada, Y., Nomura, T., Nakashima, K., & Inoue, Y. (2012). Insomnia as a risk for depression: a longitudinal epidemiologic study on a Japanese rural cohort. *The Journal of clinical psychiatry*, *73*(3), 377–83. doi:10.4088/JCP.10m06286

Oquendo, M. A., Barrera, A., Ellis, S. P., Li, S., Burke, A. K., Grunebaum, M., … Mann, J. J. (2004). Instability of symptoms in recurrent major depression: a prospective study. *The American journal of psychiatry*, *161*(2), 255–61. doi:10.1176/appi.ajp.161.2.255

Ostergaard, S. D., Jensen, S. O. W., & Bech, P. (2011). The heterogeneity of the depressive syndrome: when numbers get serious. *Acta psychiatrica Scandinavica*, *124*(6), 495–6. doi:10.1111/j.1600-0447.2011.01744.x

Pae, C. U., Tharwani, H., Marks, D. M., Masand, P. S., & Patkar, A. A. (2009). Atypical depression: A comprehensive review. *CNS drugs*, *23*(12), 1023–1037.

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*. New York, NY: Oxford University Press.

Paradis, A. D., Reinherz, H. Z., Giaconia, R. M., & Fitzmaurice, G. (2006). Major depression in the transition to adulthood: the impact of active and past depression on young adult functioning. *The Journal of nervous and mental disease*, *194*(5), 318–23. doi:10.1097/01.nmd.0000217807.56978.5b

Parker, G. B., Fink, M., Shorter, E., Taylor, M. A., Akiskal, H. S., Berrios, G., … Swartz, C. (2010). Issues for DSM-5: whither melancholia? The case for its classification as a distinct mood disorder. *American Journal of Psychiatry*, *167*(7), 745–747. doi:10.1176/appi.ajp.2010.09101525

Paykel, E. S. (2003). Life events and affective disorders. *Acta psychiatrica Scandinavica. Supplementum*, *108*(418), 61–6.

Paykel, E. S. (2008). Basic concepts of depression. *Dialogues in Clinical Neuroscience*, *10*, 279–289.

Pearl, J. (2000). *Causality:Models, Reasoning, and Inference*. Cambridge MA: Cambridge University Press.

Pearl, J. (2012). Detecting Latent Heterogeneity (Technical Report R-406). Retrieved July 18, 2013, from http://ftp.cs.ucla.edu/pub/stat_ser/r406.pdf

Perkonigg, A., Kessler, R. C., Storz, S., & Wittchen, H. U. (2000). Traumatic events and post-traumatic stress disorder in the community: prevalence, risk factors and comorbidity. *Acta psychiatrica Scandinavica*, *101*(1), 46–59.

Perlis, M. L., Giles, D. E., Buysse, D. J., Tu, X., & Kupfer, D. J. (1997). Self-reported sleep disturbance as a prodromal symptom in recurrent depression. *Journal of affective disorders*, *42*(2-3), 209–12.

Perlis, R. H., Uher, R., Perroud, N., & Fava, M. (2012). Do suicidal thoughts or behaviors recur during a second antidepressant treatment trial? *The Journal of clinical psychiatry*, *73*(11), 1439–42. doi:10.4088/JCP.12m07777

Perry, P. J. (1996). Pharmacotherapy for major depression with melancholic features: relative efficacy of tricyclic versus selective serotonin reuptake inhibitor antidepressants. *Journal of affective disorders*, *39*(1), 1–6.

Persons, J. B. (1986). The advantages of studying psychological phenomena rather than psychiatric diagnoses. *The American psychologist*, *41*(11), 1252–60.

Peterson, M. J., & Benca, R. M. (2008). Sleep in Mood Disorders. *Sleep Medicine Clinics*, *3*(2), 231–249. doi:10.1016/j.jsmc.2008.01.009

Piccinelli, M. (2000). Gender differences in depression: Critical review. *The British Journal of Psychiatry*, *177*(6), 486–492. doi:10.1192/bjp.177.6.486

Pigeon, W. R., Hegel, M., Unützer, J., Fan, M.-Y., Sateia, M. J., Lyness, J. M., … Perlis, M. L. (2008). Is insomnia a perpetuating factor for late-life depression in the IMPACT cohort? *Sleep*, *31*(4), 481–8.

Pigott, H. E., Leventhal, A. M., Alter, G. S., & Boren, J. J. (2010). Efficacy and effectiveness of antidepressants: current status of research. *Psychotherapy and psychosomatics*, *79*(5), 267–79. doi:10.1159/000318293

Pilcher, J. J., & Huffcutt, A. I. (1996). Effects of sleep deprivation on performance: a meta-analysis. *Sleep*, *19*(4), 318–26.

Pincus, H. A., Davis, W. W., & McQueen, L. E. (1999). "Subthreshold" mental disorders. A review and synthesis of studies on minor depression and other "brand names". *The British journal of psychiatry : the journal of mental science*, *174*, 288–96.

Prelow, H. M., Weaver, S. R., & Swenson, R. R. (2006). Competence, Self-esteem, and Coping Efficacy as Mediators of Ecological Risk and Depressive Symptoms in Urban African American and European American Youth. *Journal of Youth and Adolescence*, *35*(4), 506–516. doi:10.1007/s10964-006-9068-z

Prentice, D. a., & Miller, D. T. (2007). Psychological Essentialism of Human Categories. *Current Directions in Psychological Science*, *16*(4), 202–206. doi:10.1111/j.1467-8721.2007.00504.x

Psychiatric GWAS Consortium Bipolar Disorder Working Group. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature genetics*, *43*(10), 977–83. doi:10.1038/ng.943

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*(7256), 748–52. doi:10.1038/nature08185

R Development Core Team. (2008). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Radloff, L. S. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, *1*(3), 385–401. doi:10.1177/014662167700100306

Raphael, K. G., & Cloitre, M. (1994). Does mood-congruence or causal search govern recall bias? A test of life event recall. *Journal of clinical epidemiology*, *47*(5), 555–64.

Raphael, K. G., Cloitre, M., & Dohrenwend, B. P. (1991). Problems of recall and misclassification with checklist methods of measuring stressful life events. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, *10*(1), 62–74.

Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A., & Kupfer, D. J. (2013). DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses. *The American journal of psychiatry*, *170(1)*, 59–70. doi:10.1176/appi.ajp.2012.12070999

Reilly, M. C., Zbrozek, A. S., & Dukes, E. M. (1993). The Validity and Reproducibility of a Work Productivity and Activity Impairment Instrument. *PharmacoEconomics*, *4*(5), 353–365.

Reynolds, C. F., Frank, E., Houck, P. R., Mazumdar, S., Dew, M. A., Cornes, C., … Kupfer, D. J. (1997). Which elderly patients with remitted depression remain well with continued interpersonal psychotherapy after discontinuation of antidepressant medication? *The American journal of psychiatry*, *154*(7), 958–62.

Reynolds, C. F., Kupfer, D. J., Hoch, C. C., Stack, J. A., Houck, P. R., & Berman, S. R. (1986). Sleep deprivation in healthy elderly men and women: effects on mood and on sleep during recovery. *Sleep*, *9*(4), 492–501.

Roane, B. M., & Taylor, D. J. (2008). Adolescent insomnia as a risk factor for early adult depression and substance abuse. *Sleep*, *31*(10), 1351–6.

Robinson, N. S., Garber, J., & Hilsman, R. (1995). Cognitions and stress: direct and moderating effects on depressive versus externalizing symptoms during the junior high school transition. *Journal of abnormal psychology*, *104*(3), 453–63.

Rojo-Moreno, L., Livianos-Aldana, L., Cervera-Martínez, G., Dominguez-Carabantes, J. A., & Reig-Cebrian, M. J. (2002). The role of stress in the onset of depressive disorders. A controlled study in a Spanish clinical sample. *Social psychiatry and psychiatric epidemiology*, *37*(12), 592–8. doi:10.1007/s00127-002-0595-y

Rosen, G. M., Spitzer, R. L., & McHugh, P. R. (2008). Problems with the post-traumatic stress disorder diagnosis and its future in DSM V. *The British journal of psychiatry*, *192*(1), 3–4. doi:10.1192/bjp.bp.107.043083

Rosse, R., Fanous, A., Gaskins, B., & Deutsch, S. (2007). Side Effects in the Modern Psychopharmacology of Depression. *Primary psychiatry*, *14*(7).

Roy, A., Breier, A., Doran, A. R., & Pickar, D. (1985). Life events in depression. Relationship to subtypes. *Journal of affective disorders*, *9*(2), 143–8.

Ruscio, J., Zimmerman, M., McGlinchey, J. B., Chelminski, I., & Young, D. (2007). Diagnosing major depressive disorder XI: a taxometric investigation of the structure underlying DSM-IV symptoms. *The Journal of nervous and mental disease*, *195*(1), 10–9. doi:10.1097/01.nmd.0000252025.12014.c4

Rush, A. J. (2007). The varied clinical presentations of major depressive disorder. *The Journal of clinical psychiatry*, *68*(Suppl 8), 4–10.

Rush, A. J., Bernstein, I. H., Trivedi, M. H., Carmody, T. J., Wisniewski, S., Mundt, J. C., … Fava, M. (2006). An Evaluation of the Quick Inventory of Depressive Symptomatology and the Hamilton Rating Scale for Depression: A Sequenced Treatment Alternatives to Relieve Depression Trial Report. *Biological psychiatry*, *59*(6), 493–501. doi:10.1016/j.biopsych.2005.08.022

Rush, A. J., Fava, M., Wisniewski, S. R., Lavori, P. W., Trivedi, M. H., Sackeim, H. A., … Niederehe, G. (2004). Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Controlled clinical trials*, *25*(1), 119–142. doi:10.1016/S0197-2456(03)00112-0

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., … Keller, M. B. (2003). The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-Report (QIDS-SR): A Psychometric Evaluation in Patients with Chronic Major Depression. *Biological Psychiatry*, *54(5)*, 573–583. doi:10.1016/S0006-3223(03)01866-8

Rybarczyk, B., Lopez, M., Benson, R., Alsten, C., & Stepanski, E. (2002). Efficacy of two behavioral treatment programs for comorbid geriatric insomnia. *Psychology and Aging*, *17*(2), 288–298. doi:10.1037//0882-7974.17.2.288

Samkoff, J. S., & Jacques, C. H. (1991). A review of studies concerning effects of sleep deprivation and fatigue on residents' performance. *Academic medicine: journal of the Association of American Medical Colleges*, *66*(11), 687–93.

Sass, H. (2007). *Anthology of German Psychiatric Texts*. World Psychiatric Association.

Schizophrenia Psychiatric GWAS Consortium. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature genetics*, *43*(10), 969–76. doi:10.1038/ng.940

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222. doi:10.1016/j.hrmr.2008.03.003

Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, *31*(1), 43–53. doi:10.1016/j.newideapsych.2011.02.007

Sen, S., Kranzler, H. R., Didwania, A. K., Schwartz, A. C., Amarnath, S., Kolars, J. C., … Guille, C. (2013). Effects of the 2011 duty hour reforms on interns and their patients: a prospective longitudinal cohort study. *JAMA internal medicine*, *173*(8), 657–62. doi:10.1001/jamainternmed.2013.351

Sen, S., Kranzler, H. R., Krystal, J. H., Speller, H., Chan, G., Gelernter, J., & Guille, C. (2010). A prospective cohort study investigating factors associated with depression during medical internship. *Archives of general psychiatry*, *67*(6), 557–65. doi:10.1001/archgenpsychiatry.2010.41

Serretti, A., & Chiesa, A. (2009). Treatment-emergent sexual dysfunction related to antidepressants: a meta-analysis. *Journal of clinical psychopharmacology*, *29*(3), 259–66. doi:10.1097/JCP.0b013e3181a5233f

Shafer, A. B. (2006). Meta-analysis of the Factor Structures of Four Depression Questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology*, *62*(1), 123–146. doi:10.1002/jclp

Shanafelt, T. D., Bradley, K. A., Wipf, J. E., & Back, A. L. (2002). Burnout and self-reported patient care in an internal medicine residency program. *Annals of internal medicine*, *136*(5), 358–67.

Shanafelt, T., & Habermann, T. (2002). Medical residents' emotional well-being. *JAMA : the journal of the American Medical Association*, *288*(15), 1846–7.

Shi, J, Potash, J. B., Knowles, J. A., Weissman, M. M., Coryell, W., Scheftner, W. A., … Levinson, D. F. (2011). Genome-wide association study of recurrent early-onset major depressive disorder. *Molecular psychiatry*, *16*(2), 193–201. doi:10.1038/mp.2009.124

Shi, Jianxin, Levinson, D. F., Duan, J., Sanders, A. R., Zheng, Y., Pe'er, I., … Gejman, P. V. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*, *460*(7256), 753–7. doi:10.1038/nature08192

Shrout, P. E., Link, B. G., Dohrenwend, B. P., Skodol, A. E., Stueve, A., & Mirotznik, J. (1989). Characterizing life events as risk factors for depression: the role of fateful loss events. *Journal of abnormal psychology*, *98*(4), 460–7.

Sibille, E., & French, B. (2013). Biological substrates underpinning diagnosis of major depression. *The international journal of neuropsychopharmacology*, *14*, 1–17. doi:10.1017/S1461145713000436

Sjöström, N., Waern, M., & Hetta, J. (2007). Nightmares and sleep disturbances in relation to suicidality in suicide attempters. *Sleep*, *30*(1), 91–5.

Slade, T., & Andrews, G. (2005). Latent structure of depression in a community sample: a taxometric analysis. *Psychological Medicine*, *35*(4), 489–497. doi:10.1017/S0033291704003708

Solomon, A., Haaga, D. A., & Arnow, B. A. (2001). Is clinical depression distinct from subthreshold depressive symptoms? A review of the continuity issue in depression research. *The Journal of nervous and mental disease*, *189*(8), 498–506.

Solomon, D A, Keller, M. B., Leon, A. C., Mueller, T. I., Lavori, P. W., Shea, M. T., … Endicott, J. (2000). Multiple recurrences of major depressive disorder. *The American journal of psychiatry*, *157*(2), 229–33.

Solomon, David A, Leon, A. C., Coryell, W., Mueller, T. I., Posternak, M., Endicott, J., & Keller, M. B. (2008). Predicting recovery from episodes of major depression. *Journal of affective disorders*, *107*(1-3), 285–91. doi:10.1016/j.jad.2007.09.001

Spitzer, R., Kroenke, K., & Williams, J. (1999). Validation and utility of a self-report version of PRIME-MD. *JAMA*, *282*(18), 1737–1744.

Stahl, S. M., Grady, M. M., Moret, C., & Briley, M. (2005). SNRIs: their pharmacology, clinical efficacy, and tolerability in comparison with other classes of antidepressants. *CNS spectrums*, *10*(9), 732–47.

Stein, J. A., Leslie, M. B., & Nyamathi, A. (2002). Relative contributions of parent substance use and childhood maltreatment to chronic homelessness, depression, and substance abuse problems among homeless women: mediating roles of self-esteem and abuse in adulthood. *Child abuse & neglect*, *26*(10), 1011–27.

Stewart, J. W., McGrath, P. J., Fava, M., Wisniewski, S. R., Zisook, S., Cook, I., … John Rush, A. (2010). Do atypical features affect outcome in depressed outpatients treated with citalopram? *The international journal of neuropsychopharmacology*, *13*(1), 15–30. doi:10.1017/S1461145709000182

Sullivan, P F, de Geus, E. J. C., Willemsen, G., James, M. R., Smit, J. H., Zandbelt, T., … Penninx, B. W. J. H. (2009). Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Molecular psychiatry*, *14*(4), 359–75. doi:10.1038/mp.2008.125

Sullivan, Patrick F, Neale, M. C., & Kendler, K. S. (2000). Genetic Epidemiology of Major Depression : Review and Meta-Analysis. *American Journal of Psychiatry*, (October), 1552–1562.

Symister, P., & Friend, R. (2003). The influence of social support and problematic support on optimism and depression in chronic illness: a prospective study evaluating self-esteem as a mediator. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, *22*(2), 123–9.

Tansey, K. E., Guipponi, M., Perroud, N., Bondolfi, G., Domenici, E., Evans, D., … Uher, R. (2012). Genetic predictors of response to serotonergic and noradrenergic antidepressants in major depressive disorder: a genome-wide analysis of individual-level data and a meta-analysis. *PLoS medicine*, *9*(10), e1001326. doi:10.1371/journal.pmed.1001326

Taylor, S. E., Way, B. M., Welch, W. T., Hilmert, C. J., Lehman, B. J., & Eisenberger, N. I. (2006). Early family environment, current adversity, the serotonin transporter promoter polymorphism, and depressive symptomatology. *Biological psychiatry*, *60*(7), 671–6. doi:10.1016/j.biopsych.2006.04.019

Thase, M. E., Rush, A. J., Manber, R., Kornstein, S. G., Klein, D. N., Markowitz, J. C., … Keller, M. B. (2002). Differential effects of nefazodone and cognitive behavioral analysis system of psychotherapy on insomnia associated with chronic forms of major depression. *The Journal of clinical psychiatry*, *63*(6), 493–500.

The Lancet. (2012). Living with grief. *Lancet*, *379*(9816), 589. doi:10.1016/S0140-6736(12)60248-7

Trindade, E., Menon, D., Topfer, L., & Coloma, C. (1998). Adverse effects associated with selective serotonin reuptake inhibitors and tricyclic antidepressants: a meta-analysis. *Canadian Medical Association Journal*, *159*(10), 1245–1252.

Tukey, J. W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist*, *24*(2), 83–91. doi:10.1037/h0027108

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *The New England journal of medicine*, *358*(3), 252–60. doi:10.1056/NEJMsa065779

Turner, H. A., & Butler, M. (2003). Direct and Indirect Effects of Childhood Adversity on Depressive Symptoms in Young Adults. *Journal of Youth and Adolescence*, *32*(2), 89–103.

Turner, H. A., & Kopiec, K. (2006). Exposure to Interparental Conflict and Psychological Disorder Among Young Adults. *Journal of Family Issues*, *27*(2), 131–158. doi:10.1177/0192513X05280991

Tweed, D. L. (1993). Depression-related impairment: estimating concurrent and lingering effects. *Psychological Medicine*, *23*(02), 373–386. doi:10.1017/S0033291700028476

Uher, R, Farmer, A., Maier, W., Rietschel, M., Hauser, J., Marusic, A., … Aitchison, K. J. (2008). Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychological medicine*, *38*(2), 289–300. doi:10.1017/S0033291707001730

Uher, R, Perlis, R. H., Henigsberg, N., Zobel, A., Rietschel, M., Mors, O., … McGuffin, P. (2012). Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms. *Psychological medicine*, *42*(5), 967–80. doi:10.1017/S0033291711001905

Uher, Rudolf, Dernovsek, M. Z., Mors, O., Hauser, J., Souery, D., Zobel, A., … Farmer, A. (2011). Melancholic, atypical and anxious depression subtypes and outcome of treatment with escitalopram and nortriptyline. *Journal of affective disorders*, *132*(1-2), 112–20. doi:10.1016/j.jad.2011.02.014

Van Londen, L., Molenaar, R. P., Goekoop, J. G., Zwinderman, A. H., & Rooijmans, H. G. (1998). Three- to 5-year prospective follow-up of outcome in major depression. *Psychological medicine*, *28*(3), 731–5.

Vuorilehto, M., Melartin, T., & Isometsä, E. (2005). Depressive disorders in primary care: recurrent, chronic, and co-morbid. *Psychological medicine*, *35*(5), 673–82.

Wakefield, J. C. (1997). Diagnosing DSM-IV Part I: DSM-IV and the concept of disorder. *Behaviour research and therapy*, *35*(7), 633–49.

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical care*, *30*(6), 473–83.

Widaman, K. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, *28*(3), 263–311.

Williams, C. D., Taylor, T. R., Makambi, K., Harrell, J., Palmer, J. R., Rosenberg, L., & Adams-Campbell, L. L. (2007). CES-D four-factor structure is confirmed, but not invariant, in a large cohort of African American women. *Psychiatry research*, *150*(2), 173–80. doi:10.1016/j.psychres.2006.02.007

Wilson, R., Barker, M., & Brigandt, I. (2007). When Traditional Essentialism Fails: Biological Natural Kinds. *Philosophical Topics*.

Wood, A. M., Taylor, P. J., & Joseph, S. (2010). Does the CES-D measure a continuum from depression to happiness? Comparing substantive and artifactual models. *Psychiatry Research*, *177*(1-2), 120–123. doi:10.1016/j.psychres.2010.02.003

Wray, N. R., Pergadia, M. L., Blackwood, D. H. R., Penninx, B. W. J. H., Gordon, S. D., Nyholt, D. R., … Sullivan, P. F. (2012). Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Molecular psychiatry*, *17*(1), 36–48. doi:10.1038/mp.2010.109

Xie, Y., Brand, J. E., & Jann, B. (2012). Estimating Heterogeneous Treatment Effects with Observational Data. *Sociological methodology*, *42*(1), 314–347.

Zachar, P. (2002). The Practical Kinds Model as a Pragmatist Theory of Classification. *Philosophy, Psychiatry, & Psychology*, *9*(3), 219–227. doi:10.1353/ppp.2003.0051

Zachar, P., & Bartlett, S. (2002). Basic emotions and their biological substrates: A nominalistic interpretation. *Consciousness & emotion*, *2*, 189–221.

Zachar, P., & Kendler, K. (2007). Psychiatric disorders: a conceptual taxonomy. *American Journal of Psychiatry*, *164*(4), 557–565.

Zimmerman, M., Chelminski, I., McGlinchey, J. B., & Young, D. (2006). Diagnosing major depressive disorder X: can the utility of the DSM-IV symptom criteria be improved? *The Journal of nervous and mental disease*, *194*(12), 893–7. doi:10.1097/01.nmd.0000248970.50265.34

Zimmerman, M., Martinez, J. H., Friedman, M., Boerescu, D., Attiullah, N., & Toba, C. (2012). How can we use depression severity to guide treatment selection when measures of depression categorize patients differently? *The Journal of clinical psychiatry*, *73*(10), 1287–91. doi:10.4088/JCP.12m07775

Zimmerman, M., McGlinchey, J. B., Posternak, M., Friedman, M., Boerescu, D., & Attiullah, N. (2008). Remission in depressed outpatients: more than just symptom resolution? *Journal of psychiatric research*, *42*(10), 797–801. doi:10.1016/j.jpsychires.2007.09.004

Zimmerman, M., McGlinchey, J. B., Young, D., & Chelminski, I. (2006a). Diagnosing major depressive disorder I: A psychometric evaluation of the DSM-IV symptom criteria. *The Journal of nervous and mental disease*, *194*(3), 158–63. doi:10.1097/01.nmd.0000202239.20315.16

Zimmerman, M., McGlinchey, J. B., Young, D., & Chelminski, I. (2006b). Diagnosing major depressive disorder III: can some symptoms be eliminated from the diagnostic criteria? *The Journal of nervous and mental disease*, *194*(5), 313–7. doi:10.1097/01.nmd.0000217806.16329.ff

Zimmerman, M., & Spitzer, R. L. (1989). Melancholia: From DSM-III to DSM-III-R. *The American journal of psychiatry*, *146*(January), 20–28.

Zisook, S., Corruble, E., Duan, N., Iglewicz, A., Karam, E. G., Lanuoette, N., … Young, I. T. (2012). The bereavement exclusion and DSM-5. *Depression and anxiety*, *29*(5), 425–43. doi:10.1002/da.21927

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(4), 1193–8. doi:10.1073/pnas.1119675109

Appendix

List of Figures

List of Tables

List of Abbreviations

AIC	Akaike's Information Criterion
APA	American Psychiatric Association
BDI	Beck Depression Inventory
DSM	Diagnostic and Statistical Manual of Mental Disorders
CES-D	Center for Epidemiological Studies Depression Scale
CFA	Confirmatory Factor Analysis
CI	Confidence Interval
FDA	Food and Drug Administration
GAD	Generalized Anxiety Disorder
GLMM	Generalized Linear Mixed Model
GWAS	Genome-Wide Association Study
HAM-D	Hamilton Depression Rating Scale
HPC	Homeostatic Property Cluster
IVR	Interactive Voice Response System
M	Arithmetic Mean
MAOI	Monoamine Oxidase Inhibitor
MDD	Major Depressive Disorder
MDE	Major Depression Episode
MIMIC	Multiple Indicator Multiple Cause
NIMH	National Institute of Mental Health
OR	Odds Ratio
PCA	Principal Component Analysis
PHQ-9	Patient Health Questionnaire
PTSD	Posttraumatic Stress Disorder
RI	Relative Importance
SD	Standard Deviation
SEM	Structural Equation Model
SES	Socioeconomic Status
SSRI	Selective Serotonin Reuptake Inhibitor
STAR*D	Sequenced Treatment Alternatives to Relieve Depression

Curriculum Vitae

[Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.]

Publications and Talks

Publications based on this dissertation

Chapter ›5 is based on a research paper published as:

Fried, E. I., Nesse, R. M., Zivin, K., Guille. C., Sen, S. (2013). Depression is more than the sum-score of its parts: individual DSM symptoms have different risk factors. *Psychological Medicine*. Published online 02 December 2013. PuDOI: 10.1017/S0033291713002900.

Chapter ›6 will be submitted as research paper until the defense on January 29th 2014:

Fried, E. I., Nesse, R. M., Guille. C., Sen, S. The differential influence of life stress on individual symptoms of depression: a longitudinal cohort study of medical residents.

Chapter ›7 is based on a research paper currently under revision:

Fried, E. I., Nesse, R. M. (PLOS ONE, minor revision). The Impact of Individual Depressive Symptoms on Impairment of Psychosocial Functioning.

Talks based on this dissertation

1. Fried E. I., et al. (2013). Depression is more than the sum-score of its parts: covert heterogeneity of depression symptoms. Invited lecture at the University of Amsterdam, the Netherlands.
2. Fried E. I., et al. (2013). Breaking the wall of depression diagnosis. Talk at International Conference on Future Breakthroughs in Science and Society, Berlin, Germany.
3. Fried E. I., et al. (2013). Heterogeneity of individual depressive symptoms in a longitudinal study of medical residents. Talk at the European Congress of Psychology, Stockholm, Sweden.
4. Fried E. I., et al. (2013). Depression is more than the sum-score of its parts: covert heterogeneity of depressive symptoms. Invited lecture at the Centre for Behaviour & Evolution, Newcastle, UK.
5. Fried, E. I., et al. (2012). Why depression symptoms matter. Invited lecture at the University of Michigan, Ann Arbor, USA.
6. Fried, E. I., et al. (2012). Symptomatic heterogeneity in depression: A longitudinal study of depressive symptoms in medical residents. Talk at the International Congress of Psychology, Cape Town, South-Africa.

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel erarbeitet und verfasst habe. Diese Arbeit in Gänze oder einzelne Teile waren nicht Gegenstand eines früheren Promotionsvorhabens.

Berlin, 15. Dezember 2013

Eiko Fried